

-199-

orf75-1		ATLADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD	
		190 200 210 220 230 240	
5	orf75a.pep	250 260 270 280 290	
		EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX	
orf75-1		EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX	
		250 260 270 280 290	
<b>Homology with a predicted ORF from <i>N.gonorrhoeae</i></b>			
ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from <i>N.gonorrhoeae</i> :			
15	orf75.pep	MFVFQTAFXMFOKHLQKASDSVVGGLYVVATPIGNLADITLRALAVLQKA----AEDTR	56
	orf75ng	MSVFQTAFMFQKHLQKASDSVVGGLYVVATPIGNLADITLRALAVLQKADIICAEDTR	60
20	orf75.pep	VTAQLLSAYGIQGLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR	116
	orf75ng	VTAQLLSAYGIQGLVSVREHNERQMADKIVGYLSDGLVVAQVSDAGTPAVCDPGAKLAR	120
25	orf75.pep	RVREAGFKVVPVVGAXAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPV	176
	orf75ng	RVREAGFKVVPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPV	180
30	orf75.pep	MFETPHRIGAAALADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALSADGDQSRGEM	236
	orf75ng	MFETPHRIGATLADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM	240
35	orf75.pep	VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYD	288
	orf75ng	VLVLYPAQDEKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>:

35	1	MSVFQTAFMF	FQKHLQKASD	SVVGGLYVV	ATPIGNLADI	TLRALAVLQK
	51	ADIICAEDTR	VTAQLLSAYG	IQGLVSVRE	HNERQMADKV	IGFLSDGLVV
40	101	AQVSDAGTPA	VCDPGAKLAR	RVREAGFKVV	PVVGASAVMA	ALSAGVAES
	151	DFYFNGFVPP	KSGERRKLFA	KWVRAAFPVV	MFETPHRIGA	TLADMAELFP
	201	ERRRLMLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE
	251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK
	301	*				

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

45	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCTTGC
	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTACAG	GCAGGTGGT
50	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
	251	TCCTTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
	301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCCTCG	TGGGCGCAAG	CGCGGTAATG	GCGGCGTTGA
55	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTT	ATTTCACCGG	TTTGTACCG
	451	CCGAAATCGG	GCGAAGCTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGGC
	501	ATTCTCTGTC	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
	551	CCGATATGGC	GGAATTGTTC	CCGAACGCC	GTCTGATGCT	GGCGCGCGAA
60	601	ATCACGAAAA	CGTTTGAAC	GTCTTAAGC	GGCACGGTTG	GGGAAATTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGCGGAG	ATGGTGTG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
	751	CAAAATGCGA	TGAAATCCT	TGCGGCCGAG	CTGCCGACCA	AGCAGGCGGC
60	801	GGAGCTTGCC	GCCAAGATTA	CAGGTGAGGG	CAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	GTGGAATAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>:

-200-

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQGRVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP  
 151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF PERRLMLARE  
 201 ITKTFTFELS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:

		10	20	30	40	50	60
10	orf75-1.pep	MFQKHLQKASDSVVGGTLYV	VATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY				
	orf75ng-1	MFQKHLQKASDSVVGGTLYV	VATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY				
		10	20	30	40	50	60
15	orf75-1.pep	GIQKLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKV					
	orf75ng-1	GIQKLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKV					
		70	80	90	100	110	120
20	orf75-1.pep	VPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLF	FAKWVRAAFPVVMFETPHRIG				
	orf75ng-1	VPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLF	FAKWVRAAFPVVMFETPHRIG				
		130	140	150	160	170	180
25	orf75-1.pep	ATLADMAELFPERRLMLAREITKTFTFSLSGTVGEIQTAL	AADGNQSRGEMVLVLYPAQD				
	orf75ng-1	ATLADMAELFPERRLMLAREITKTFTFSLSGTVGEIQTAL	AADGNQSRGEMVLVLYPAQD				
		190	200	210	220	230	240
30	orf75-1.pep	EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX					
	orf75ng-1	EKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNKX					
		250	260	270	280	290	
35	orf75-1.pep	EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX					
	orf75ng-1	EKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNKX					
		250	260	270	280	290	

Furthermore, ORG75ng-1 shows significant homology to a hypothetical *E.coli* protein:

sp|P45528|YRAL\_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)  
 40 >gi|606086 (U18997) ORF\_f286 [Escherichia coli]  
 >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region [Escherichia coli] Length = 286  
 Score = 218 bits (550), Expect = 3e-56  
 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)  
 45 Query: 4 KHLQKASDSVVGGTLYVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63  
 K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI  
 Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLAQVDLIAEDTRHTGLLLQHFGIN 59  
 50 Query: 64 GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123  
 RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+  
 Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119  
 55 Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLF

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 35

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

1 ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTGGCAGG
51 TTTTGCGGCA GC.AAAGCAC CCGAAATCGA CCCGGCTTTG .....
//
651 ..... ..GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
10 701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AAACGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

```

1 MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL .....
//
15 201 ..... ELVRNQLEQG LRQEKARLKI DALLEENGVK
251 P*

```

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

```

1 ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTGGCAGG
51 TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
20 101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTACAAAC
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAGACGAGCT
25 351 GCACAAGTTT TACGAACAGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC
551 AGTTTGCCGC GATGAATCGG GGCAGCGTTA CCCGCGATCC GGTCAAATTG
30 601 GCGCAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AAACGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

```

35 1 MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
51 KPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNRK KIAEASFYAE
101 EYVRFLESE TVSEDELHKF YEQQIRMIKL QVVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
201 GERYLFLKLS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENGK
40 251 KP*

```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) from strain A of *N. meningitidis*:

```

45 orf76.pep      10      20      30
      MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL
      |||||
orf76a      MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
      10      20      30      40      50      60
      //
                        70      80      90

```

-202-

```

orf76.pep                                XELVRNQLEQGLRQEKARLKIDALLEENGVKPX
                                         |||||:|||||
orf76a      DVTRDPVKLGERYLFLKLEVGKNPDAQPFELVRNQLEQGLRQEKARLKIDALLEENGVKPX
              200      210      220      230      240      250

```

5 The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

```

1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTGGCAGG
51  TTTTGGCGCA GCCAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAAC
10  201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
351 GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
15  451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCGGAACG ACGAGCAGGC
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCAGAGCCG CTGGCTTCGC
551 AGTTTGCAGC GATGAATCGG GCGCAGCTTA CCCGCGATCC GGTCAAATTG
601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAACAA GGTTCGAGAC
20  701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCA TTTTGGAAGA AAACGGTGTC
751 AAACCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 300>:

```

1  MKQKKTAAGV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
25  51  KPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNRF KIAEASFYAE
101 EYVRFLESE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
201 GERYLFLKLS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENGV
251 KP*

```

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

```

30  orf76a.pep      10      20      30      40      50      60
      MKQKKTAAGVIAAMLAGFAAAKAPALVDTLVAQIMQADRHAESQKPDGQAIRND
      orf76-1      10      20      30      40      50      60
      MKQKKTAAGVIAAMLAGFAAAKAPALVDTLVAQIMQADRHAESQKPDGQAIRND
35  orf76a.pep      70      80      90      100     110     120
      AVRRLQTLVLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSESALRQF
      orf76-1      70      80      90      100     110     120
      AVRRLQTLVLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSEDELHKF
40  orf76a.pep      130     140     150     160     170     180
      YERQIRMIKLQQVSFATEEEARQAQQLLLKLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      orf76-1      130     140     150     160     170     180
      YEQQIRMIKLQQVSFATEEEARQAQQLLLKLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
45  orf76a.pep      190     200     210     220     230     240
      LASQFAAMNRGDVTRDPVKLGERYLFLKLEVGKNPDAQPFELVRNQLEQGLRQEKARLK
      orf76-1      190     200     210     220     230     240
      LASQFAAMNRGDVTRDPVKLGERYLFLKLEVGKNPDAQPFELVRNQLEQGLRQEKARLK
50  orf76a.pep      250
      IDALLEENGVKPX
      orf76-1      250
      IDALLEENGVKPX

```

## 60 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:



-203-

```

orf76.pep      MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL      30
|||||
orf76ng        MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND  60
//
5 orf76.pep      ELVRNQLEQGLRQEKAARKLIDALLEENGVKP  251
|||||
orf76ng        VTRNPVKLGERYLFLKLGAVGKNPDAPFELVRNQLEQGLRQEKAARKLIDALLEENGVKP  251

```

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

```

1  ATGAAACAGA AAAAGACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
10 51 TTTTGGCGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AGACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTGCAAAC
201 TTTGGAAGTT TTGAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
15 301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
351 GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
501 GLTFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR GDVTRNPVKL
20 551 agtttgCCGG TATGAACCGT GCGGACGTTA CCCGCAATCC GGTCAAATTG
601 GGCGAACGCT ATTACCTGTT CAAACTCGGC GCGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTTTGAGGC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA Aaacggtgtc
751 AaacCGTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 302>:

```

1  MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
51  RPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNR F KIAEASFYAE
101 EYVRFLESE TVSESALRQF YERQIRMIKL QVVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR GDVTRNPVKL
30 201 GERYLFLKLG AVGKNPDAPF FELVRNQLEQ GLRQEKAARK IDALLEENGV
251 KP*

```

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

```

10 20 30 40 50 60
35 orf76-1.pep MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND
orf76ng      MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND
10 20 30 40 50 60
40 orf76-1.pep AVRRLQTLV LKNRALKEGLDKDKDVQNRFKIAEASFYAE EYVRFLESETVSEDELHKF
orf76ng      AVRRLQTLV LKNRALKEGLDKDKDVQNRFKIAEASFYAE EYVRFLESETVSESALRQF
70 80 90 100 110 120
45 orf76-1.pep YEQQIRMIKLQVVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
orf76ng      YERQIRMIKLQVVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
130 140 150 160 170 180
50 orf76-1.pep LASQFAAMNRGDVTRDPVKLGERYLFLKLGAVGKNPDAPFELVRNQLEQGLRQEKAARKL
orf76ng      LASQFAGMNRGDVTRNPVKLGERYLFLKLGAVGKNPDAPFELVRNQLEQGLRQEKAARKL
190 200 210 220 230 240
55 orf76-1.pep IDALLEENGVKPX
orf76ng      IDALLEENGVKPX
250
60 orf76ng      IDALLEENGVKPX
250

```

Furthermore, ORF76ng shows significant homology to a *B.subtilis* export protein precursor:

```

sp|P24327|PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|98227|pir||S15269
33K lipoprotein - Bacillus subtilis >gi|39782 (X57271) 33kDa lipoprotein
[Bacillus subtilis]
>gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
5 >gi|2633331|gnl|PID|e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]
Length = 292
Score = 50.4 bits (118), Expect = 1e-05
Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)

10 Query: 70 VLKNRALKEGLDK-----DKDVQNRFKIAEASF-----YAEYVRFLERSETVSE 114
      VL ++ LDK DK++ N+ K + Y ++Y++ + E +++
Sbjct: 53 VLTQLVQEKVLDKKYKVSDEIDNKLKEYKTQLGQYTALEKQYKDYLKEQVKYELLTQ 112

15 Query: 115 SA-----LRQFYERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPN 163
      A +++++E I+ + A ++ A + ++ L KG FE L K Y
Sbjct: 113 KAAKDNIKVTDADIKEYWEGELGKIRASHILVADKKTAEVEKKLKKGEKFDLAKEYST 172

Query: 164 DEQAFDG-----FIMAQQLPEPLASQFAAMNRGDVTRDPVKLGERYYLFKLSEVGKNPDA 218
      D A G F Q+ E + + G+V+ DPVK Y++ K +E D
20 Sbjct: 173 DSSASKGGDLGWFAKEGQMDLTFSKAAFKLKTGEVS-DPVKTQYGYHIIKKTEERGKYDD 231

Query: 219 QPFELVRNQLEQGLRQEKA 237
      EL LEQ L A
25 Sbjct: 232 MKKELKSEVLEQKLNDA 250

```

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

### 35 Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

```

1 ATGAAAAAAT CTTTCCTTAC GCTTGTTCTG TATTCGTCTT TACTTACCGC
51 CAGCGAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGCG
101 GCAAAAATTG CGGAAACGTT TGCGCTGACA TTTGTGATTG CTGCGCTGTA
40 151 TCTGTTTGCG CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG
201 CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG
251 ATGACG.... //
1201 ..... CAAACCGTAT TCGAGCAGCT GCAAAAGACT CCTGACGGCA
45 1251 ACTGGCTGTT TGCCTATACC TCCGATCATG GCCAGTATGT TCGCCAAGAT
1301 ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCGCTAGT
1351 GTTGTAACAG CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
1401 CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTTCAACGTT CCTGATTCAC
1451 ACGTTGGGCT ACGATAATGCC GGTTTCAGGT TGTCCGGAAG GCTCGGTAAC
50 1501 GGGCAACCTG ATTACGGGTG ATGCAGGCAG CTTGAACATT CGCGACGGCA
1551 AGGCGGAATA TGTATTATCCG CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

```

1 MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAKIAETF ALTFVIAALY

```

```

51  LFARNKVTRL  LIAVFFAFSI  IANNVHYADY  QSWMT.....
                                     //
401  ...QTVFEQL  QKTPDGNWLF  AYTSDHGQYV  RQDIYNQGTV  QPDSYLVPLV
451  LYSPDKAVQQ  AANQAFAPCE  IAFHQQLSTF  LIHTLGMDVP  VSGCREGSVT
501  GNIITGDAGS  INIRDGKAEE  VYPO*

```

10

15

20

25

30

35

1	ATGAAAAAAT	CTTTCCTTAC	GCTTGTCTGT	TATTCGTCTT	TACTTACCGC
51	CAGCGAAATT	GCCTATCGCT	TTGTATTTTG	GATTGAAACC	TTACCGCGCG
101	CAAAAATTTG	GGAAACGTTT	GCGCTGACAT	TTGTGATTGC	TGCGCTGTAT
151	CTGTTTGCGG	GTTATAAGGT	GACGCGTTTG	TTGATTGCGG	TGTTTTTTGC
201	GTTCAGCATT	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGA
251	TGACGGGGAT	CAATTATTGG	CTGATGCTGA	AAGAGGTTAC	CGAAGTCGGG
301	ACGCGCGGTG	CGTCGATGTT	GGATAAGTTT	TGGCTGCCTG	TGTTGTGGGG
351	CGCTTTGGAA	GTCATGTTGT	TTTGACGCGT	TGCCAAGTTC	CGCGCTAAGA
401	CGCATTTTTC	TGCCGATATA	CTGTTTGCCT	TCCTAATGCT	GATGATTTTC
451	GTGCGTCCGT	TGCACACGAA	ACAAGAGCAC	GGTATTTCGC	CCAAACCGAC
501	ATACGAGCCG	ATCAAAGCCA	ATTATTTTCAG	CTTCGGTTAT	TTTGTCCGAG
551	GCGTGTTCGC	GTATCAGTTG	TTTGATTTAA	CGAGGATTCC	CGCCTTTAAG
601	CAGCCTGCTC	CAAGCAAAAT	CGGGCAGGGC	AGTGTTCAAA	ATATCGTCCT
651	GATTATGGGC	GAAAGCGAAA	GCGCGCGCGA	TTTGAAGCTG	TTTGGCTACG
701	GACGCGAAAC	TTCCGCGTTT	TTAACCCCGG	TGTCGCAAGC	CGATTTTAAAG
751	CCGATTGTGA	AACAAAGTTA	TTCCGCAAGC	TTTATGACTG	CAGTGCCTCT
801	GCCCAGTTTT	TTCAATGCGA	TACCGCACGC	CAACGGCTTG	GAACAAATCA
851	GCGCGCGGGA	TACCAATATG	TTCCGCGCTC	CCAAAGAGCA	GGCGTATGAA
901	ACGTATTTTT	ACAGCGCGCA	GCGGGAACAA	GAGATGGCGA	TTTTGAACTT
951	AATCGGTAAG	AAATGGATAG	ACCAATCTGAT	TCAGCCGACG	CAACTTGCGT
1001	ACGGCAACGG	CGACAATATG	CCCGATGAGA	AGCTGCTGCC	GTTGTTTCGAC
1051	AAAATCAATT	TGCAGCAGGG	CAAGCATTTT	ATCGTGTTGC	ACCAACCGCG
1101	TTGCGCACCC	CCATACCGGC	CAATTGTTTAT	GCCTCAAGAT	AAAGTATTCG
1151	GCGAAGCCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAAACCGAC
1201	CAAATGATTC	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
1251	CTGGCTGTTT	GCCATATACCT	CCGATCATGG	CCAGTATGTT	CGCGCAAGATA
1301	TCTACAATCA	AGGCACGGTG	CAGCCCCGAC	GCTATCTCGT	CGCGCTTAGTG
1351	TTGTACAGCC	CGGATAAGGC	CGTGAACAAG	GCTGCCAACCC	AGGCTTTTTGCG
1401	GCCTTGCAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTTCACA
1451	CGTTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCCGGAAGG	CTCGGTAAACG
1501	GGCAACCTGA	TTACGGGTGA	TTCAGGCGAG	TTGAACATTTC	GCGACGGCAA
1551	GGCGGAATAT	GTTTATCCGC	AATGA		

40

45

50

1	MKKSFLTTLV	YSSLLTASEI	AYRFVFGIET	LPAAKIAETF	ALTFTVIAALY
51	LFARYKVTRL	LIAEVFAFSI	IANNVHYAVY	QSWMTGINYW	LMLEKMTVEVG
101	SAGASMLDKL	WLPLVLGVLE	VMLECSFLAKV	RRKTHFSADI	LFAEFLVLE
151	VRSFDTKQEH	GISPKPTYSR	IKANYFSFGY	FGVRVLPLYQ	FDLSRIPAFK
201	QPAPSKIGQG	SVQNIIVLIMG	ESESAAHLKL	FGYGRETSFF	LTRLSQLADFK
251	PIYKQSYSAG	EMTAVSLPSF	FNAIPLHANG	EQTSGGDNTM	FRLAKEQGYE
301	TFYFSAQAEN	EMAILNLIGK	KWIDHLIQPT	QLQYNGNDNM	PRDEKLPLFD
351	KINLQQKGHF	IVLHQRGSHA	PYGALLQPQD	KVFGEDIVD	KYDNTIHKTD
401	QMIQTVEEQL	QKQPDGNWLF	AYTSDHGQXV	RQDLYNQGTV	QPDSYLVPLV
451	LYSPDKAVQQ	AAANQAFAPCE	IAFHQQLSTF	LIHTLGMDP	VSGCREGSVT
501	GNLITDGAGS	LNIRDGKAEY	YVPO*		

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF81 shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) from strain A of *N. meningitidis*:

55

60

```

              10      20      30      40      50      60
orf81.pep  MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALTFVIAALYLFARNKVTRL
            |||||:::| ||||| ||||| : ||||| ||||| ||||| ||||| ||||| : |||||
orf81a     MKKSFLVFLYLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL
              10      20      30      40      50      60

              70      80

```

-206-

5	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT
	orf81a	LIAVFFAFSIIANNVHYAVYQSWITGINYWLMLKEITEVGGAGASMLDKLWLPALWGVLE
		70 80 90 100 110 120
		//
	orf81.pep	QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD
	orf81a	IPHANGLEQISGGDIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD
10		280 290 300 310 320 330
	orf81.pep	IYNQGTVPDSYLVPLVLYSPDKAVQQAAQAFAPCEIAFHQQQLSTFLIHTLGYDMPVSG
15	orf81a	IYNQGTVPDSYLVPLVLYSPDKAVQQAAQAFAPCEIAFHQQQLSTFLIHTLGYDMPVSG
		340 350 360 370 380 390
	orf81.pep	CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
20	orf81a	CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
		400 410 420

The complete length ORF81a nucleotide sequence <SEQ ID 307> is:

25	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTTCTG	TATTCGTCCC	TACTTACTGC
	51	CAGCGAAATT	GCTTATCGCT	TTGTATTCGG	AATTGAAACC	TTACCGGCTG
	101	CAAAAATGGC	AGAAACGTTT	GCGCTGACAT	TTGTGATTGC	TGCGCTGTAT
	151	CTGTTTGCGC	GTTATAAGGC	AACGCGTTTG	TTGATTGCGG	TGTTTTTCGC
	201	GTTCAGCATT	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGA
	251	TAACGGGCAT	TAATTATTGG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
30	301	GGCGCAGGGG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CGTTGTGGGG
	351	CTGTTGGAA	GTCATGTTGT	TTTGCAGCCT	TGCCAAGTTC	CGCCGTAAGA
	401	CGCATTTTTC	TGCCGATATA	CTGTTTGCC	TCCTAATGCT	GATGATTTTC
	451	GTGCGTTCGT	TCGACACGAA	ACAAGAACAC	GGTATTTTCG	CCAAACCGAC
35	501	ATACAGCCGC	ATCAAAGCCA	ATTATTTTTC	CTTCGGTTAT	TTTGTCCGAC
	551	GCGTGTGGCC	GTATCAGTTG	TTTGATTTAA	GCAAGATTCC	TGTGTTCAAA
	601	CAGCTGCTC	CAAGCAGAAT	CGGCAAGGC	AGTATTCAAA	ATATCGTCCT
	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATTG	TTTGGCTACG
	701	GGCGCGAAAC	TTCCGCGTTT	TTGACCCAGC	TTTCGCAAGC	CGATTTTAA
	751	CCGATTGTGA	AACAAAGTTA	TTCCGCGAGC	TTTATGACGG	CAGTATCCCT
40	801	GCCAGTTTTC	TTTAACGTCA	TACCGCATGC	CAACGGCTTG	GAACAAATCA
	851	GCGGCGGCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAACCGGAC
	901	CAAAATGATT	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
	951	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTT	CGCCAAGATA
45	1001	TCTACAATCA	AGGCACGGTG	CAGCCCGACA	GCTATCTCGT	GCCGCTGGTG
	1051	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	GCTGCCAACC	AGGCTTTTGC
	1101	GCCTTGCGAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTC	CTGATTACAC
	1151	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCGCGAAGG	CTCGGTAACG
	1201	GGCAACCTGA	TTACGGGTGA	TGCAGGCAGC	TTGAACATTC	GCGACGGCAA
	1251	GGCGGAATAT	GTTTATCCGC	AATGA		

50 This encodes a protein having amino acid sequence <SEQ ID 308>:

	1	MKKSFLVFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAETF	ALTFVIAALY
	51	LFARYKATRL	LIAVFFAFSI	IANNVHYAVY	QSWITGINYW	LMLKEITEVG
	101	GAGASMLDKL	WLPALWGVLE	VMLFCSLAKE	RRKTHFSADI	LFAFLMLMIF
55	151	VRSFDTKQEH	GISPKPTYSR	IKANYFSFGY	FVGRVLPYQL	FDLSKIPVFK
	201	QPAPSRIGQG	SIQNIIVLIMG	ESESAAHLKL	FGYGRETSFP	LTQLSQADFK
	251	PIVKQSYSAG	FMTAVSLPSF	FNVIPHANGL	EQISGGDIVD	KYDNTIHKTD
	301	QMIQTVFEQL	QKQPDGNWLF	AYTSDHGQYV	RQDIYNQGTV	QPDSYLVPLV
	351	LYSPDKAVQQ	AANQAFAPCE	IAFHQQQLSTF	LIHTLGYDMP	VSGCREGSVT
	401	GNLITGDAGS	LNIRDGKAEY	VYPQ*		

60 ORF81a and ORF81-1 show 77.9% identity in 524 aa overlap:

		10	20	30	40	50	60
	orf81a.pep	MKKSFLVFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL					
		::					
65	orf81-1	MKKSFLTLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFAITFVIAALYLFARYKVTRL					
		10	20	30	40	50	60

-207-

		70	80	90	100	110	120
	orf81a.pep	LI	AVFFAFSII	ANNVHYAVYQSWITG	INYLMLKEITEVGGAGASMLDKLWLPALWGVLE		
5	orf81-1	LI	AVFFAFSII	ANNVHYAVYQSWMTG	INYLMLKEVTEVGSAGASMLDKLWLPVLWGVLE		
		70	80	90	100	110	120
	orf81a.pep	130	140	150	160	170	180
10	orf81-1	VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSR	KANYFSFGY				
		130	140	150	160	170	180
	orf81a.pep	190	200	210	220	230	240
15	orf81-1	FVGRVLPYQLFDLSKIPVFKQPAPSRIGQGSIQNIVLIMGESESA	AHLKLF	FGYGRETSPF			
		190	200	210	220	230	240
20	orf81a.pep	250	260	270	280		
	orf81-1	LTQLSQADFKPIVKQSYSAGFMTAVSLPSFFNVIPHANGLEQISGGD					
25		250	260	270	280	290	300
	orf81a.pep	-----					
30	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPL	FDKINIQQKHF				
		310	320	330	340	350	360
	orf81a.pep	-----	290	300	310	320	
35	orf81-1	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVFEQ	LQKQPDGNWLF				
		370	380	390	400	410	420
	orf81a.pep	330	340	350	360	370	380
40	orf81-1	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
	orf81a.pep	390	400	410	420		
45	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKA	EYVYPQX				
		490	500	510	520		
50	<u>Homology with a predicted ORF from <i>N.gonorrhoeae</i></u>						
	The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from <i>N. gonorrhoeae</i> of the						
	N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:						
	orf81.pep	MKKSFLT	LVLYSSLLTASEIAYPLELGIETLPAAKIAETFALT	FVIAALYL	FARNKVTRL	60	
55	orf81ng	MKKS	LFVLFYSSLLTASEIAYRFVFGIETLPAAKMAETFALT	FMIAALYL	FARYKASRL	60	
	orf81.pep	LI	AVFFAFSII	ANNVHYADYQSWMT		85	
60	orf81ng	LI	AVFFAFSMT	ANNVHYAVYQSWMTG	INYLMLKEVTEVGSAGASMLDKLWLPALWGVAE	120	
	orf81.pep				QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD	433	
	orf81ng	ALLQ	PQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD			433	
65	orf81.pep	IYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG				493	
	orf81ng	IYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG				493	

5	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTTCTG	TATTCATCCC	TACTTACCGC
	51	CAGCGAAATC	GCCTATCGCT	TTGTATTCCG	AATTGAAACC	TTACCGGCTG
	101	CAAAAAATGGC	GGAAACGTTT	GCGCTGACAT	TTATGATTGC	TGCGCTGTAT
	151	CTGTTTCGCG	GTTATAAGGC	TTCGCGGCTG	CTGATTGCGG	TGTTTTTCGC
10	201	GTTCCAGCATG	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAAGCTGGA
	251	TGACGGGTAT	TAAC TATTGG	CTGATGCTGA	AAGAGTTTAC	CGAAGTCGGC
	301	AGCGCGGGCG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CTTTGTGGGG
	351	CGTGGCGGAA	GTCATGTTGT	TTTGCAGCCT	TGCCAAGTTT	CGCCGTAAAG
15	401	CGCATTTTTTC	GTCCGATATA	CTGTTTGCCT	TCCTAATGCT	GATGATTTTT
	451	GTGCGTTCGT	TCGACACGAA	ACAAGAGCAC	GGTATTTCGC	CCAAACCGAC
	501	ATACAGCCCG	ATCAAAGCCA	ATTATTTTTCAG	CTTCGGTTAT	TTTGTGCGGC
	551	GCGTGTGTGCC	GTATCAGTTG	TTTGATTTAA	CGAAGATCCC	TGTGTTCAAA
20	601	CAGCCTGCTC	CAAGCAAAAT	CGGGCAAGGC	AGTATTCAAA	ATATCGTCCT
	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATTG	TTTGGTTACG
	701	GGCGCGAAAC	TTCCGCGTTT	TTAACC CGCG	TGTCGCAAGC	CAGATTTTAA
	751	CCGATTGTGA	AACAAAGTTA	TTCCGCAAGC	TTTATGACGG	CGAGTATCCCT
25	801	GCCCAGTTTC	TTTAACTGCA	TACCGCACGC	CAACGGCTTG	GAACAAATCA
	851	GCGGCGGCGA	TACCAATATG	TTCCGCGCTCG	CCAAAGAGCA	GGGCTATGAA
	901	ACGTATTTTT	ACAGTGGCCC	GGCTGAAAA	CAAAATGGCA	TTTTGAACTT
	951	AATCGTAAAG	AAATGGATAG	ACCATCTGAT	TCCAGCGACG	CAACTTGGCT
30	1001	ACGGCAACGG	CGACAATATG	CCCATGAGA	AGCTGCTGCC	GTTGTTTCGAC
	1051	AAAATCAATT	TGCAGCAGGG	CAGGCATTTT	ATCGTGTGCG	ACCAACGCGG
	1101	TTCCGACGCG	CCATACGCGC	CATTGTTGCA	GCCTCAAGAT	AAAGTATTCG
	1151	GCGAAGCCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAAACCTGAC
35	1201	CAAATGATTC	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
	1251	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTG	CGCCAAAGATA
	1301	TCTACAATCA	AGGCACGGTG	CAGCCCAGCA	GCTATATTGT	GCCTCTGGTT
	1351	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	GCTGCCAACC	AGGCTTTTGC
40	1401	GCCTTTCGAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTACAA
	1451	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCGCGAAGG	CTCGGTAAAC
	1501	GGCAACCTGA	TTACGGGCGA	TGCAGGCAGC	TTGAACATTG	GCAACGGCAA
	1551	GGCGGAATAT	GTTTATCCGC	AATAA		

	1	<u>MKKS</u> <u>SLFV</u> <u>LF</u> <u>FL</u>	<u>YSSLLT</u> <u>ASEI</u>	<u>AYRFV</u> <u>FGIET</u>	<u>LPAAKMA</u> <u>ETF</u>	<u>ALTFMIA</u> <u>AALY</u>
40	51	<u>LFARY</u> <u>KASRL</u>	<u>LTAVFF</u> <u>AFSM</u>	<u>TANNV</u> <u>HYAVY</u>	<u>QSWMT</u> <u>GINYW</u>	<u>LMLEK</u> <u>VEV</u> <u>Q</u>
	101	<u>SAGAS</u> <u>MDKL</u>	<u>WLPAL</u> <u>WGVAE</u>	<u>VMLFCS</u> <u>IAKF</u>	<u>RRKTH</u> <u>ESADI</u>	<u>LF AFLM</u> <u>LMIF</u>
	151	<u>VRSFD</u> <u>TQEH</u>	<u>GLSPK</u> <u>PTYSR</u>	<u>IKANY</u> <u>FSFGY</u>	<u>FGVGR</u> <u>LYPQL</u>	<u>FDLSK</u> <u>IPVFK</u>
	201	<u>QFAPS</u> <u>KIGQG</u>	<u>SIQNI</u> <u>VILIM</u>	<u>ESESA</u> <u>AHLKL</u>	<u>FGYGR</u> <u>ETSPF</u>	<u>LTRL</u> <u>SQADF</u> <u>K</u>
	251	<u>PIVKQ</u> <u>SYSAG</u>	<u>FMTAV</u> <u>SLPSF</u>	<u>FNVI</u> <u>PHANGL</u>	<u>EQISG</u> <u>GDTNM</u>	<u>FRLAKE</u> <u>QGYE</u>
45	301	<u>TYFY</u> <u>SAQAN</u>	<u>QMAIL</u> <u>NLIGK</u>	<u>KWIDH</u> <u>LQPT</u>	<u>QLVYG</u> <u>GDNDM</u>	<u>PDEK</u> <u>LLPFD</u>
	351	<u>KINL</u> <u>QQRHF</u>	<u>IVLH</u> <u>QRGSHA</u>	<u>PYGAL</u> <u>LQPD</u>	<u>KVFGE</u> <u>ADIVD</u>	<u>KYDNT</u> <u>IHKTD</u>
	401	<u>QMIQT</u> <u>VFEQL</u>	<u>QKQPD</u> <u>GNWLF</u>	<u>AYTSD</u> <u>HQQYV</u>	<u>RQDIY</u> <u>NGQTV</u>	<u>QPDSY</u> <u>IPLV</u>
	451	<u>LYSPD</u> <u>KAVQ</u>	<u>AAHQ</u> <u>AFAPCE</u>	<u>IAFHQ</u> <u>QLSTF</u>	<u>LIHTL</u> <u>GYDMP</u>	<u>VSGCR</u> <u>EGSVT</u>
	501	<u>GNLIT</u> <u>GDA</u> <u>GS</u>	<u>LNIRN</u> <u>GKAEY</u>	<u>VYPO</u> *		

50		10	20	30	40	50	60
	orf81ng-1.pep	MKKS	LFVLFLYSS	LLTASEIAYRFV	FGIETLPAAKMAET	FALTFMIAALYLF	FARYKASRL
			: : :			: :	
	orf81-1	MKKS	FLLTLVLYSS	LLTASEIAYRFV	FGIETLPAAKIAET	FALTFVIAALYLF	FARYKVTRL
55		10	20	30	40	50	60
	orf81ng-1.pep	LI	AVFFAFS	MIANNVHYAVY	QSWMTG	INYLMLKEVTE	VGSAGASMLDKLWLPALW
	orf81-1	LI	AVFFAFS	IANNVHYAVY	QSWMTG	INYLMLKEVTE	VGSAGASMLDKLWLPVLWG
60		70	80	90	100	110	120
	orf81ng-1.pep	LI	AVFFAFS	MIANNVHYAVY	QSWMTG	INYLMLKEVTE	VGSAGASMLDKLWLPALW
	orf81-1	LI	AVFFAFS	IANNVHYAVY	QSWMTG	INYLMLKEVTE	VGSAGASMLDKLWLPVLWG
65		70	80	90	100	110	120
	orf81ng-1.pep	VML	FCSLAKFR	RKTHFSADIL	FAFLMLMIF	VRSFDTKQEH	GISP
	orf81-1	VML	FCSLAKFR	RKTHFSADIL	FAFLMLMIF	VRSFDTKQEH	GISP
		130	140	150	160	170	180
	orf81ng-1.pep	VML	FCSLAKFR	RKTHFSADIL	FAFLMLMIF	VRSFDTKQEH	GISP
	orf81-1	VML	FCSLAKFR	RKTHFSADIL	FAFLMLMIF	VRSFDTKQEH	GISP

-209-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf81ng-1.pep	FVGRVLPYQLFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESA AHLKLFYGRGRETSPF					
	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESA AHLKLFYGRGRETSPF					
		190	200	210	220	230	240
		250	260	270	280	290	300
10	orf81ng-1.pep	LTRL SQADFKPIVKQSY SAGFMTAVSLP SFFNVIP HANGLEQISGGDTNMFRLAKEQGYE					
	orf81-1	LTRL SQADFKPIVKQSY SAGFMTAVSLP SFFNAIPHANGLEQISGGDTNMFRLAKEQGYE					
		250	260	270	280	290	300
		310	320	330	340	350	360
15	orf81ng-1.pep	TYFYSAQAENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGRHF					
	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF					
		310	320	330	340	350	360
20		370	380	390	400	410	420
	orf81ng-1.pep	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTQDQMIQTVFEQLQKQPDGNWLF					
	orf81-1	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTQDQMIQTVFEQLQKQPDGNWLF					
25		370	380	390	400	410	420
		430	440	450	460	470	480
	orf81ng-1.pep	AYTSDHGQYVRQDIYNQGTVPQDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
30	orf81-1	AYTSDHGQYVRQDIYNQGTVPQDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
		490	500	510	520		
35	orf81ng-1.pep	LIHTLGYDMPVSGCREG SVTGNLITGDAGSLNIRNGKAEYVYPQX					
	orf81-1	LIHTLGYDMPVSGCREG SVTGNLITGDAGSLNIRDGKAEYVYPQX					
		490	500	510	520		

Furthermore, ORF81ng shows significant homology to an *E.coli* OMP:

40	gi 1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)
45	Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFARYKAS--RLLIAVFFAFSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLL+A F + A ++ ++Y Sbjct: 29 VFGITNLVASSGAHMQRLFFVLTLVVKRISSPLRLLVAAPFVL-LTAADMSISLY- 86
50	Query: 82 SWMT-----GINYWLMLKEVTEVGSAGASMLDKLWLPALWGVAEVMFLFCSLAKFRRKT 134 SW T G ++ + EV A ML ++ P L A + L + Sbjct: 87 SWCTFGTTTFNDGFAISVLQSDPDEV----AKMLG-MYSPYLCAFAFLSLLFLAVIIKYDV 141
55	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISPKPTYSRIKAN--YFSFGYFVG 183 + L+L++ S D K ++ SP SR ++F+ YF Sbjct: 142 SLPTKKVTGILLLVISGSFLSACQFAYKDAKNKNAFSPYILASRFATYTPFFNLNYFAL 201
60	Query: 184 RVLPHYQ--LFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESA AHLKLFYGRGRETSPFL 241 +Q L + +P F+ + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVFPYQL----SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQV 257
65	Query: 242 TRLSQADFKPIVKQSY SAGFMTAVSLP---SFFNVIP HANGLEQISGGDTNMFRLAKEQG 298 +Q + Q+ S TA+S+P + +V+ H I N+ +A + G Sbjct: 258 E--AQRKQIKLENQAISGAPYTALSVPLSLTADSVLSH-----DIHNPYDNIINMANQAG 310
70	Query: 299 YETIFYSAQA---ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355 ++T++ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q Sbjct: 311 FQTFWLSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLPLHLSQALQQ 359
	Query: 356 --QGRHFIVLHQRGSHAPYGALLQPDQKVFGEADIVDK-YDNTIHKTQDQMIQTVFEQLQK 412 Q + IVLH GSH P + VF D D YDN+IH TD ++ VFE L+ Sbjct: 360 NTQQKKLIVLHLNGSHEPACSAYPQSSAVFQPDQDQACYNDSIHYTDSLGLQGVFELLK- 418

Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQG--TVQPD SYIVPL-VLYSP 454  
           D          Y +DHG          ++++Y G          +Y VP+ + YSP  
 Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAYHVPFIWYSP 464

5

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

15  
20  
25  
30  
35  
40  
45  
50  
55

```

1   ...ACCCTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGCG GCACACTGAC
51  CGGCATACTC GCCCaCGGCG GCGGCAAACG CT'TTGCCGTC GAACAAGAAC
101 TCGTCGCGCG AT'CGTCCCGC GCGGCCGTCA AAGAAATGGA TT'TGTCCGCC
151 y'TAAAAGGAC GCAAAGCCGC CyTTTACGTC TCCGTTATGG GCGACCAAGG
201 TT'CGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
251 GCGGCTACCA CAACAACCCC GAAAGTGCCA CCAATACAG CTACCCCGCC
301 TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCGTAACCAC
351 TTCCACATCG CTTTTGAACG CCCCCGCCGC CGyCyTGACG AAAACAGCG
401 GACGCAAAGG CGAACGcTCC GCGGACTGT CCGTCAACGG CACGGGCGAC
451 TACCGCAACG AAACCTTGCT CGCCAACCCC CGGACGTTT CTT'CTTGAC
501 CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCGC
551 CCGrATACGC CGACACCGAC GTATTCTGTA CCGTCGACGT A...
```

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

25  
30  
35  
40  
45  
50  
55

```

1   ..TLLLFIPVL TXCGTLTGIL AHGGGKRFAV EQELVAASSR AAVKEMDLA
51  LKGRKAAXYV SVMGDQSGN ISGGRYSIDA LIRGGYHNNP ESATQYSYPA
101 YDTTATTKSD ALSSVTTSTS LLNAPAAXLT KNSGRKGRS AGLSVNGTGD
151 YRNETLLANP RDVSFLTNLI QTVFYLRGIE VVPPXYADTD VFVTVDV...
```

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90

```

1   ATGAAAACCC TGCTCCTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC
51  ACTGACCGGC ATACCCGCCC ACGGCGGCGG CAAACGCTTT GCCGTGGAAC
101 AAGAACTCGT CGCCGCATCG TCCCGCGCGG CCGTCAAAGA AATGGATTTG
151 TCCGCCCTAA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA
201 CCAAGGTTTC GGCAACATAA GCGGCGGACG CTACTCTATC GACGACTGA
251 TACGCGGCGG CTACCACAAC AACCCTGAAA GTGCCACCCA ATACAGCTAC
301 CCCGCCTACG ACATACCGC CACCACCAA TCCGACGCGC TCTCCAGCGT
351 AACCCTTCC ACATCGCTTT TGAACGCCCC CGCCGCCGCC CTGACGAAAA
401 ACAGCGGACG CAAAGGCGAA CGTCCGCGG GACTGTCCGT CAACGGCACC
451 GGCGACTACC GCAACGAAAC CCTGCTCGCC AACCCTCGCG ACGTTTCCTT
501 CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG
551 TACCGCCCGA ATACGCGGAC ACCGACGTAT TCGTAACCGT CGACGTATTG
601 GGCACCGTCC GCAGCCGTAC CGAACTGCAC CTCTACAACG CCGAAACCTT
651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTTGACCGC GACAGCCGGA
701 AACTGCTGAT TACCCCTAAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA
751 CAATACGCCC TTTGGACCGG CCCTTACAAA GTCAGCAAAA CCGTCAAAGC
801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATTACCCCC TACGGCGACA
851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAAACCC
901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

50  
55

```

1   MKTLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
51  SALKGRKAAL YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY
101 PAYDTTATTK SDALSSVTTTS TSLNAPAAA LTKNSGRKGE RSAGLSVNGT
151 GDYRNETLLA NRPDVSFLTN LIQTVFYLRG IEVVPPEYAD TDVFVTVDVV
201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLITPK TAAYESQYQE
251 QYALWTGPYK VSKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNGKKP
```



Homology with a predicted ORF from *N.meningitidis* (strain A)

5 *meningitidis:*

		10	20	30	40	50	
	orf83.pep	<u>TLLLFIFPLVLTXC</u> GTLTGILAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAX					
		:					
10	orf83a	<u>MKTLLXLIPLVLTA</u> CGLTGTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL					
		10	20	30	40	50	60
		60	70	80	90	100	110
	orf83.pep	YVSVMGDQGSGNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKS DALSSVTTS					
15	orf83a	YVSVMGDQGSGNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKS DALSSVTTS					
		70	80	90	100	110	120
		120	130	140	150	160	170
	orf83.pep	TSLLNAPAAXLTKNSGRKGERSAGLSVNGTDGYRNETLLANPRDVSF LTNLIQT VFYL RG					
20	orf83a	TSLLNAPAAALTKNSGRKGERSAGLSVNGTDGYRNETLLANPRDVSF LTNLIQT VFYL RG					
		130	140	150	160	170	180
		180	190				
	orf83.pep	IEVVPPXYADTDVF VTV DV					
25	orf83a	IEVVPEYADTDVF VTV DV FGT VR SRT EL HLY NA ET L KA Q TK LE Y FA VD RD SR K LL I AP K					
		190	200	210	220	230	240

30	1	ATGAAACCC	TGCTCNCCT	CATCCCCCTC	GTCCTCACAG	CCTGCGGCAC
	51	ACTGACCGGC	ATACCCGCC	ACGGCGGCGG	CAAACGCTTT	GCCGTCGAAC
	101	AAGAACTCGT	CGCCGCATCG	TCCCGCGCCG	CCGTCAAAGA	AATGGACTTG
	151	TCCGCCCTGA	AAGACGCAA	AGCCGCCCTT	TACGTCTCGG	TTATGGGCGA
35	201	CCAAGGTTG	GGCAACATAA	GCGGCGGACG	CTACTCTATC	GACGCACTGA
	251	TACGCGGCGG	CTACCACAAC	AACCCCGAAA	GTGCCACCCA	ATACAGCTAC
	301	CCCGCCTACG	ACACTACCGC	CACCACCAA	TCCGACGCGC	TCTCCAGCGT
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCCC	CGCCGCGCCG	CTGACGAAAA
40	401	ACAGCGGACG	CAAAGGCGAA	CGCTCCGCGG	GACTGTCCGT	CAACGGCACG
	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCCCGCGG	ACGTTTCTCTT
	501	CCTGACCAAC	CTCATCCAA	CCGTCTTCTA	CCTGCGCGGC	ATCGGAAGTC
	551	TACCGCCCGA	ATACGCCGAC	ACCGACGTAT	TCGTAACCGT	CGACGTATTG
45	601	GGCACCCTCC	GCAGCCGCAC	CGAACTGCAC	CTCTACAACG	CCGAAACCCT
	651	TAAAGCCCAA	ACCAAGCTCG	AATATTTCGC	CGTTGACCGC	GACAGCCGGA
	701	AACTGCTGAT	TGCCCTTAA	ACGCGCGCCT	ACGAATCCCA	ATACCAAGAA
	751	CAATACGCC	TCTGGATGGG	ACCTTACAGC	GTCGGCAAAA	CCGTCAAAGC
	801	CTCAGACCGC	CTGATGGTCG	ATTTCTCCGA	CATCACCCCC	TACGGCGACA
	851	CAACCCGCCA	AAACCGTCCC	GACTTCAAAC	AAAACAACCG	TAAAAAACCC
	901	GATGTCGGCA	ACGAAGTCAT	CCGCGCGCCG	AAAGGAGGAT	AA

50	1	MKTLLXLIPL	VLTACGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	YVSVMGDQGS	GNISGGGRYSI	DALIRSGYHN	NPEATQYSY
	101	PAYDTTATTK	NDALSSVTS	TSLLNAPAAA	LTKNSGRKGE	RSAGLSVNGT
	151	GQYRNETLLA	SPRDVSFLT	LIQTVFYLRG	IEVVPPEYAD	TDVFTVTVDF
55	201	GTVRSRTELH	LYNAETLKAQ	TKLEYFAVDR	DSRKLLIAPK	TAAYESQYQE
	251	QYALWMGPYS	VGKTVKASDR	LMVDFSDITP	YGDTTAQNRP	DFKQNGGKKP
	301	DVGNEVIRRR	KGG*			

orf83a.pep                    10                    20                    30                    40                    50                    60  
MKTLLXLIPLVLTACGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL

10

20

25

30

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N. gonorrhoeae*:

40

45

50

55

60

65

-213-

5  
 651 TAAAGCCCAA ACCAAGCTCG AATATTTTCG CGTCGACCGC GACAGCCGGA  
 701 AACTGCTGAT TGCCCCCTAAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA  
 751 CAATACGCC CTTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC  
 801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATCACCCCC TACGGCGACA  
 851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCCC  
 901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

10  
 1 MKTLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL  
 51 SALKGRKAAL YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPDSATRYSY  
 101 PAYDTTATTK SDALSGVTTs TSLNAPAAA LTKNNGRKGE RSAGLSVNGT  
 151 GDYRNETLLA NPRDVSFLTN LIQTVFYLRG IEVVPPEYAD TDVFTVVDVF  
 201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE  
 251 QYALWMGPYS VGKTVKASDR LMVDFSDITP YGDTTAQNR PDKQNNNGKNP  
 301 DVGNEVIRRR KGG\*

15 ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

20  
 orf83-1.pep 10 20 30 40 50 60  
 MKTLLLLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL  
 orf83ng MKTLLLLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL  
 25  
 orf83-1.pep 70 80 90 100 110 120  
 YVSVMGDQGS GNISGGRYSIDALIRGGYHNNPESATQSYSPAYDTTATTKSDALSSVTTs  
 orf83ng YVSVMGDQGS GNISGGRYSIDALIRGGYHNNPDSATRYSPAYDTTATTKSDALSGVTTs  
 30  
 orf83-1.pep 130 140 150 160 170 180  
 TSLNAPAAALTKNSGRKGER SAGLSVNGTGDYRNETLLANPRDVSFLTNLIQTVFYLRG  
 orf83ng TSLNAPAAALTKNNGRKGER SAGLSVNGTGDYRNETLLANPRDVSFLTNLIQTVFYLRG  
 35  
 orf83-1.pep 190 200 210 220 230 240  
 IEVVPPEYADTDVFTVVDVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLITPK  
 orf83ng IEVVPPEYADTDVFTVVDVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK  
 40  
 orf83-1.pep 250 260 270 280 290 300  
 TAAYESQYQE QYALWGPYKVS KTVKASDR LMVDFSDITPYGDTTAQNR PDKQNNGKGP  
 orf83ng TAAYESQYQE QYALWMPYS VGKTVKASDR LMVDFSDITPYGDTTAQNR PDKQNNNGKNP  
 45  
 orf83-1.pep 310  
 DVGNEVIRRRKGGX  
 orf83ng DVGNEVIRRRKGGX  
 50  
 310

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 38**

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

      1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT
5     51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
      101  AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAGGCTT GAAAATACCG
      151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
      201  GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
      251  TCGGGTCTAT TGTCAATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
10    301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
      351  ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCCT AAGCTTCTAG
      401  ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTCAAAC
      451  AAGATGGGTA TGCGTACGCT TTTAGAAATGG AAAATATGCG CGGACGATCC
      501  aGGAAGAACG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
15    551  AAGTTTATGA CTTGTATsrr TmmGCGGAAG TTCATACCGT AAATAAGGTC
      601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
      651  CGTGTTTGTC GGCCTGTCCCT ATAAAATGTT GagCaGTTAC GGAAAAAAC
      701  aGGAAGAACG CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
      751  CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
20    801  AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGrAAGC AAGCcgatTT
      851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
      901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
      951  gaAAGAAGTG ACGGaGTTGA TGTGcgAgG aCTATGTaAA AAacGGCTTG
25   1001  CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGGCAGGAAG TTCAGCAAAG
      1051  CGCGCAGCAA CATTCGGACA GGGCGgCAAG TTGCCACATT GGGCGGAAAA
      1101  CCGTAGCAGA ACCTAATTGTA CGATAATTGG GAAGAACGCG GGAAACCGTT
      1151  TGAAGGAATC GGaCGGGGGC GTGTCGGAT CGGCAAACTG A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

      1  MAEICLITGT PGSGKTLKMV SMMANDEMFK PDEKAIRKRV FTNIKGLKIP
30    51  HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR
      101  SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
      151  KMGMRLLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYX XAEVHTVNVK
      201  KRSKWFTTLP VIVLLIPVTV GLSYKMLSSY GKKQEEPAAQ ESAATEQQAV
      251  LPDKTEGEPV NNGNLTADMF VPTLSEKPKS KPIYNGVRQV RTFEYIAGCI
35    301  EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGEVQQS
      351  AQQHSdraQV ATLGKPKXQN LMYDNWEERG KPFEIGGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

      1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT
40    51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
      101  ACGGCATACG CCGTAAAGTA TTTACGAACA TAAAGGCTT GAAAATACCG
      151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
      201  GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
      251  TCGGGTCTAT TGTCAATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
45    301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
      351  ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCCT AAGCTTCTAG
      401  ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTCAAAC
      451  AAGATGGGTA TGCGTACGCT TTTAGAAATGG AAAATATGCG CGGACGATCC
      501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
50    551  AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
      601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
      651  CGTGTTTGTC GGCCTGTCCCT ATAAAATGTT GAGCAGTTAC GGAAAAAAC
      701  AGGAAGAACG CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
      751  CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
55    801  AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGAAAGC AAGCGGATTT
      851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
      901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCATCAAG GGACGGCATT
      951  GAAAGAAGTG ACGGAGTTGA TGTGCAAGGA CTATGTAAAA AACGGCTTGC
60   1001  CGTTTAACCC ATACAAAGAA GAAAGCCAAG GGCAGGAAGT TCAGCAAAGC
      1051  GCGCAGCAAC ATTCGGACAG GCGCAAGTT GCCACATTGG GCGGAAACC
      1101  GTAGCAGAAC CTAATGTACG ATAATTGGGA AGAACGCGGG AAACCGTTTG
      1151  AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP  
 51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR  
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN  
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNVK  
 201 KRSKWFYTLF VIVLLIPVFLV GLSYKMLSSY GKKQEEPAAQ ESAATEQQAV  
 251 LPDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTFEYIAGCI  
 301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGEVQQS  
 351 AQQHSRAQV ATLGKGP\*QN LMYDNWEERG KPFEGIGGGV VGSAN\*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N. meningitidis*:

15	orf84.pep	MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENKIRRKV FTNIKGLKIP HTYIETDAKK
	orf84a	MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK
20	orf84.pep	LPKSTDEQLSAHDMYEWIKK PENIGSIVIV DEAQDVWPARSAGSKIPENV QWLNTHRHQG
	orf84a	LPKSTDEQLSAHDMYEWIKK PENIGSIVIV DEAQDVWPARSAGSKIPENV QWLNTHRHQG
25	orf84.pep	IDIFVLTQGP KLLDQNLRTLVRKHYHIASN KMGMRITLLEWKICADDPVKM ASSAFSSIYT
	orf84a	IDIFVLTQGS KLLDQNLRTLVRKHYHIASN KMGMRITLLEWKICADDPVKM ASSAFSSIYT
30	orf84.pep	LDKKVYDLYXXAEVHTVNVKVRKSKWFYTLFVIVLLIPVFLVGLSYKMLSSY GKKQEEPAAQ
	orf84a	LDKKVYDLYESA EVHTVNVKVRKSKWFYTLFVIVLLIPVFLVGLSYKMLSSY GKKQEEPAAQ
35	orf84.pep	ESAATEQQAVLPDKTEGEPV NNGNLTADMFVPTLSEKPKSKPIYNGVRQV RTFEYIAGCI
	orf84a	ESAATEHQAVFQDKTEGEPV NNGNLTADMFVPTLSEKPKSKPIYNGVRQV RTFEYIAGCV
40	orf84.pep	EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKEESQGEVQQSAQ QHSRAQV
	orf84a	EGGRTGCTCY SHQGTALKEITKEMCKDYARNGLPFNPYKEESQGRDVQQSEQHSDRPQV
45	orf84.pep	ATLGKGPXQNL MYDNWEERG KPFEGIGGGV VGSANX
	orf84a	ATLGKGPWQNL MYDNWQERG KPFEGIGGGV VGSANX

The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

55 1 ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT  
 51 AAAAAATGGTT TCCATGATGG CAAACGATGA AATGTTTAAG CCGGATGAAA  
 101 ACGGCATACG CCGTAAAGTA TTTACGAACA TCAAAGGCTT GAAGATACCG  
 151 CACACCTACA TAGAAACGGA CGCGAAAAAG CTGCCGAAAT CGACAGATGA  
 201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA  
 251 TCGGGTCTAT TGTGATTGTA GATGAAGCTC AAGACGTATG GCCGACACGC  
 301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG  
 351 ACATCAGGGC ATTGATATAT TTGTTTGAC TCAAGGCTCT AAGCTTCTAG  
 401 ATCAAAATCT TAGAAGCCTT GTACGGAAC ATTACCACAT CGCTTCAAC  
 451 AAGATGGGTA TGCGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC

-216-

501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA  
 551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC  
 601 AAGCGGTCAA AATGGTTTAA TACTCTGCCA GTAATAATAT TGCTGATGCC  
 651 CGTTTTTGTG GGCCTGTCCCT ATAAAAATGTT AAGTAGTTAT GGAAAAAAAC  
 701 AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA TCAGGCAGTA  
 751 TTTCAGGATA AAACAGAAGG CGAGCCGGTA AACAACGGTA ACCTTACCGC  
 801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT  
 851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTGTA  
 901 GAAGGCGGAA GAACCGGATG CACATGCTAT TCGCATCAAG GGACGGCATT  
 951 GAAAGAAATT ACAAAGGAAA TGTGCAAGGA TTACGCAAGA AACGGATTGC  
 1001 CGTTTAAACC ATATAAGAA GAAAGCCAAG GGCGGGATGT CCAGCAAAGT  
 1051 GAGCAGCACC ATTCGGACAG ACCGCAAGTT GCCACGTTGG GCGGAAAGCC  
 1101 GTGGCAAAAT CTTATGTATG ATAATTGGCA GGAGCGCGGA AAACCGTTTG  
 1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

15 This encodes a protein having amino acid sequence <SEQ ID 324>:

1 MAEICLITGT PGSGKTLKLV SMANDEMFK PDENGIRRKV FTNIKGLKIP  
 51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR  
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTTQS KLLDQNLRTL VRKHYHIASN  
 151 KMGMRLLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNVK  
 201 KRKWFYTLV VIILLIPVEV GLSYKMLSSY GKKQEEPAQ ESAATEHQAV  
 251 FQDKTEGEPV NNGNLTADMF VPTLSEKPEK KPIYNGVRQV RTFEYIAGCV  
 301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS  
 351 EQHHSRDPQV ATLGKPPWQN LMYDNWQERG KPFEGIGGVV VGSAN\*

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

25 orf84a.pep 10 20 30 40 50 60  
 MAEICLITGT PGSGKTLKLV SMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK  
 orf84-1 MAEICLITGT PGSGKTLKLV SMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK  
 30 10 20 30 40 50 60  
 orf84a.pep 70 80 90 100 110 120  
 LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG  
 orf84-1 LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG  
 35 70 80 90 100 110 120  
 orf84a.pep 130 140 150 160 170 180  
 IDIFVLTTQS KLLDQNLRTL VRKHYHIASN KMGMRLLLEW KICADDPVKM ASSAFSSIYT  
 orf84-1 IDIFVLTTQS KLLDQNLRTL VRKHYHIASN KMGMRLLLEW KICADDPVKM ASSAFSSIYT  
 40 130 140 150 160 170 180  
 orf84a.pep 190 200 210 220 230 240  
 LDKKVYDLYE SAEVHTVNVK KRKWFYTLV VIILLIPVEV GLSYKMLSSY GKKQEEPAQ  
 orf84-1 LDKKVYDLYE SAEVHTVNVK KRKWFYTLV VIILLIPVEV GLSYKMLSSY GKKQEEPAQ  
 45 190 200 210 220 230 240  
 orf84a.pep 250 260 270 280 290 300  
 ESAATEHQAV FQDKTEGEPV NNGNLTADMF VPTLSEKPEK KPIYNGVRQV RTFEYIAGCV  
 orf84-1 ESAATEHQAV FQDKTEGEPV NNGNLTADMF VPTLSEKPEK KPIYNGVRQV RTFEYIAGCV  
 50 250 260 270 280 290 300  
 orf84a.pep 310 320 330 340 350 360  
 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS EQHHSRDPQV  
 orf84-1 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS EQHHSRDPQV  
 55 310 320 330 340 350 360  
 orf84a.pep 370 380 390  
 ATLGKPPWQN LMYDNWQERG KPFEGIGGVV VGSANX  
 orf84-1 ATLGKPPWQN LMYDNWQERG KPFEGIGGVV VGSANX  
 60 370 380 390  
 orf84a.pep  
 orf84-1  
 65 370 380 390

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N.gonorrhoeae*:

5	orf84.pep	MAEICLITGTPGSGKTLKVMSSMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK	60
	orf84ng	MAEICLITGTPGSGKTLKVMSSMANDEMFKPDENGVRKRVFTNIKGLKIPHTHIETDAKK	60
10	orf84.pep	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
	orf84ng	LPKSTDEQLSAHDMYEWIKKPENVGAIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
15	orf84.pep	IDIFVLTQGPKLDDQNLRLTVLRKHVHIAASNKMGMRTLLEWKICADDPVKMASSAFSSIYT	180
	orf84ng	IDIFVLTQGPKLDDQNLRLTVLRKHVHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIYT	180
20	orf84.pep	LDKKVYDLYXXAEVHTVNKVKRSKWFYTLPVIVLLIPVFGVLSYKMLSSYGKKQEEPAAQ	240
	orf84ng	LDKKVYDLYESAEIHTVNKVKRSKWFYALPVIIILLIPLFVGLSYKMLGSYGKKQEEPAAQ	240
25	orf84.pep	ESAATEQQAVLPDKTEGEPVNNGNLTADMFVPTLSEKXPXSKPIYNGVRQVRTFEYIAGCI	300
	orf84ng	ESAATEQQAVLPDKTEGESVNNGNLTADMFVPTLPEKPESKPIYNGVRQVRTFEYIAGCI	300
30	orf84.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQQSAQQHSDRAQV	360
	orf84ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQQSAQQHSDRAQV	360
35	orf84.pep	ATLGKKPXQNLMYDNWEERGKPFEGIGGGVVG SAN 395	
	orf84ng	ATLGKKPQQNLMYDNWEERGKPFEGIGGGVVG SAN 395	

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

	1	ATGGCAGAAA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTCA	GGAAAAACATT
	51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAG	CCAGATGAAA
35	101	ACGGCGTACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGTTT	GAAGATACCG
	151	CACACCCACA	TAGAAACAGA	CGCAAAGAAG	CTGCCGAAAT	CAACCGATGA
	201	ACAGCTTTTCG	GCGCATGATA	TGTATGAATG	GATCAAGAAG	CCTGAAAacg
	251	tcggcgCAAT	CGTTATTGTC	GATGAGGCGC	AAGACGTATG	GCCCGCACGC
	301	TccgCAGGTT	CGAAAATCCC	CGAAAACGTC	CAATGGCTGA	ACACACACAG
40	351	GCATCAGGGC	ATAGATATAT	TTGTATTGAC	ACAAGGTCCT	AAACTCTTAG
	401	ATCAGAACTT	GCGAACATTG	GTTAAAAGAC	ATTACCACAT	TGCGGCCAAC
	451	AAAATGGGTT	TGCGTACCCT	GCTTGAATGG	AAAGTATGCG	CGGATGACCC
	501	GATAAAAATG	GCATCAAGTG	CATTTTCCAG	TATCTACACA	CTGGATAAAA
45	551	AAGTTTATGA	CTTGTACGAA	TCCGCAGAAA	TTCACACGGT	AAACAAAGTC
	601	AAGCGTTCAA	AATGGTTTTA	TGCATTGCCC	GTCATCATAT	TATTGATTCC
	651	GCTATTTGTC	GGTTTGTCTT	ACAAAATGTT	GGGCAGTTAC	GGAAAAAAC
	701	AGGAAGAACC	CGCAGCACAA	GAATCGGCGG	CAACAGAACA	GCAGGCAGTA
50	751	CTTCCGGATA	AAACAGAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACGGC
	801	AGATATGTTT	GTTCCGACAT	TGCCCGAAAA	ACCCGAAAGC	AAGCCGATTT
	851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA
	901	GAAGGCGGAA	GAACCGGATG	CACCTGCTAT	TCGCATCAAG	GGACGGCATT
55	951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGGCTTGC
	1001	CGTTTAACCC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC
	1051	GCGCAGCAAC	ATTCGGACAG	GGCGCAAGTT	GCCACCTTGG	GCGGAAAACC
	1101	GCAGCAGAAC	CTAATGTACG	ACAATTGGGA	AGAACGCGGG	AAACCGTTTG
	1151	AAGGAATCGG	CGGGGGCGTG	GTCGGATCGG	CAAACCTGA	

This encodes a protein having amino acid sequence <SEQ ID 326>:

	1	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGVRKRV	FTNIKGLKIP
	51	HTHIETDAKK	LPKSTDEQLS	AHDMYEWIKK	PENVGAIVIV	DEAQDVWPAR
60	101	SAGSKIPENV	QWLNTHRHQG	IDIFVLTQGP	KLLDQNLRTL	VKRHYHIAAN
	151	KMGLRTLLEW	KVCADDPVKM	ASSAFSSIYT	LDKKVYDLYE	SAEIHTVNKV
	201	KRSKWYFALP	VIILLIPLFV	GLSYKMLGSY	GKKQEEPAAQ	ESAATEQQAV
	251	LPDKTEGESV	NNGNLTADMF	VPTLPEKPES	KPIYNGVRQV	RTFEYIAGCI
	301	EGGRTGCTCY	SHQGTALKEV	TELMCKDYVK	NGLPFNPYKE	ESQGQEVQQS
	351	AQQHSDRAQV	ATLGKKPQQN	LMYDNWEERG	KPFEGIGGGV	VGSAN*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

		10	20	30	40	50	60
	orf84-1.pep	MAEIC	LTGTPGSGKTLK	MVSMMA	NDEMFKPDENG	IRRKVFTN	IKGLKIPHTYIETDAKK
5	orf84ng	MAEIC	LTGTPGSGKTLK	MVSMMA	NDEMFKPDENG	VRRKVFTN	IKGLKIPHTIETDAKK
		10	20	30	40	50	60
	orf84-1.pep	70	80	90	100	110	120
10	orf84ng	LPKSTDEQLSAHDMYEWIKK	PENIGSIVIVDEA	QDVWPARSAGS	KIPENVQWL	NTHRHQG	
		70	80	90	100	110	120
	orf84-1.pep	130	140	150	160	170	180
15	orf84ng	IDIFVLTQGP	KLLDQNLRLTLV	KRHYHIA	NKMGMLRTL	LEWKVCADD	PVKMASSAFSSIYT
		130	140	150	160	170	180
	orf84-1.pep	190	200	210	220	230	240
20	orf84ng	LDKKVYDLY	ESAEHTVNKVK	RSKWFYALP	VIILLIPLF	VGLSYKMLGS	YGGKQEEPAQAQ
		190	200	210	220	230	240
	orf84-1.pep	250	260	270	280	290	300
25	orf84ng	ESAATEQQAV	LPDKTEGEPV	NNGNLTADM	FVPTLSEK	PESKPIYNG	VRQVRTFEYIAGCI
		250	260	270	280	290	300
	orf84-1.pep	310	320	330	340	350	360
30	orf84ng	EGGRTGCAC	YSHQGTALKE	VTLMCKDYV	KNGLPFNPY	KEESQGGQ	EVQQAQHS
		310	320	330	340	350	360
	orf84-1.pep	370	380	390			
35	orf84ng	ATLGGKPK	QNLMYDNWE	ERGKPFEG	GIGGGVGS	SANX	
		370	380	390			
	orf84-1.pep	370	380	390			
40	orf84ng	ATLGGKPK	QNLMYDNWE	ERGKPFEG	GIGGGVGS	SANX	
		370	380	390			

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 39

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

	1	GTGGTTTTC	TGAATGCCGA	CAACGGGATA	TGCGTTCAGG	ACTTGCCTTT
50	51	TGAAGTCAAA	CTGAAAAAAT	TCCATATCGA	TTTTTACAAT	ACGGGTATGC
	101	CGCGTGATTT	CGCCAGCGAT	ATTGAAGTGA	CGGACAAGGC	AACCGGTGAG
	151	AAACTCGAGC	GCACCATCCG	CGTGAACCAT	CCTTTGACCT	TGCACGGCAT
	201	CACGATTAT	CAGGCGAGTT	TGCGGACGG	CGGTTCCGAT	TGACATTCA
	251	AGGCGTGGAA	TTTGGGTGAT	GCTTCGCGCG	AGCCTGTCGT	GTTGAAGGCA
55	301	ACATCCATAC	ACCAGTTTCC	GTTGGAAATT	GGCAAACACA	AATATCGTCT
	351	TGAGTTCGAT	CAGTTCACTT	CTATGAATGT	GGAGGACATG	AGCGAGGGCG
	401	CGGAACGGGA	AAAAAGCCTG	AAATCCACGC	TGCCCGATGT	CCGCGCGGTT
	451	ACTCAGGAAG	GTCACAAATA	CACCAAT...	.....	.....TACCG
	501	TATCCGTGAT	GCGCCAGGCC	AGGCGGTCGA	ATATAAAAC	TATATGCTGC
60	551	CGGTTTTCGA	GGAACAGGAT	TATTTTTCGA	TTACCGGCAC	GCGCAGCGC.



5  
10

```

601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
651 AGCGGACACC TTTATGGCAT TGCCTGAGTT TTTGAAAGAT GGGGAAGGGC
701 GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC
751 GAACAATTCA TGCTGGCTGC GGAACACACG CTGAACATCT TTGCACAAA
801 AGGCTATTG GGAATGGACG AATTATTAC GTCCAATATC CCGAAAGAGC
851 AGCAGGATAA GATGCAGGGC TATTCTACG AAATGCTTTA CGGCGTGATG
901 AACGCTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA
951 GCAGGATGAA GCGCGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA
1001 CGGCTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT
1051 TCCGAGGTGC GTTCGTCGGG TTTGCAGATG ACCCGTTCCC C.GGTCCGCT
1101 TTTGGTCTAT CTC...
```

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

15  
20

```

1 MVFLNADNGI LVQDLPEFVK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
51 KLERTIRVNH PLTLHGITIY QASFADGGSD LTFKAWNLDG ASREPVVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLPDVRVAV
151 TQEGHKYTNX XXXXXYRIRD APGQAVEYKN YMLPVLQEQD YFWITGTRSX
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRRKRXVAD ATKGAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKEQQDKMQG YFYEMLYGVM
301 NAALDETXTR YGLEWQQQDE ARNRELLHSM DAYTGLTEYP APMLLQLDGF
351 SEVRSSGLQM TRSXGPELLVY L...
```

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

25  
30  
35  
40  
45  
50  
55  
60

```

1 ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGGTTCCG
51 TTTTTCAGC TCCATGCGCT TTGCAGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTGGGATC GTTTTGGGCG CAGATTTTTG GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGCTGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCAAGTTG TGCTGATTC GCAATGTGCC GCCGTCTCTG
301 GCGGAAATGA AGTCTTTTCG GGAAGAGTT AAAGAAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA
401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGGTCCG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTTCCTG GCGGGGTTGA
551 TATGCAGTAA CCTGCTGTTG AAAGTGGGTA TGCTGACCGG TCGGATTGTT
601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCGTCC AATCTCTCAT TTAGGGGCAA CGTCAATATP TCCGAGGGGC
701 AGAGTGCCTT TGTGGTTTTT CTGAATGCCG ACAACGGGAT ATTGGTTCAG
751 GAGTGCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCGTGATT TCGCCAGCGA TATTGAAGTG ACGGACAAGG
851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA
951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCG
1001 TGTTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAT TGGCAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTCAC TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CCGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TCGCGCAGGG CAGGCGGTCG AATATAAAAA
1251 CTATATGCTG CCGGTTTTCG AGGAACAGGA TTATTTTGG ATTACCGGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
1401 TGGGGAAGGG CGCAACGTC TGGTTGCCGA CGCAACCAA GGCGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CCGAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATGGAC GAATTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACGATAT
1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTGCAGAT GACCCGTTCC
1801 CCGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GGCGTGGGTA TTGTTTTACG
1901 ACGGCAAAA CCGTTTGGC ATGTCTTCGG CCCGAGCGA ACGGATTGTT
1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
2001 CTTGAATCAT GACTGA
```

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

65

```

1 MSKSRSPPL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
```

-220-

5  
10

```

101 REMKSFREKV KEKSLAAMRH SSLDVKIAP EVAKRYLEVQ GFQGKTINRE
151 DGSVLIAAKK GTMNKWDYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
201 PDNQAVYAKD FKPEISILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITIYQAS FADGGSDLTF KAWNLDGASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PVLQEODYFW ITGTRSGLQQ QYRWLRIPLD
451 KQLKADTFMA LREFLKDGEK RKRLVADATK GAPAEIREQF MLAAENTLNI
501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGVMAA LDETIRRYGL
551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTS
601 PGALLVYLG S VLLVLGTVLM FYVREKRAWV LFS DGKIRFA MSSARSERDL
651 QKEFPKHVES LQRLGKDLNH D*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

15 ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

```

20 orf88.pep          MVFLNADNGILVQDLPFEVKLKKFHIDFYN
    :|||||
orf88a      AKDFKPESILGASNLSFRGNVNISEGQSADVFLNADNGILVQDLPFEVKLKKFHIDFYN
            210      220      230      240      250      260

25 orf88.pep          40      50      60      70      80      90
    TGMPRDFASDIEVTDKATGEKLERTIRVNHPLTLHGITIIYQASFADGGSDLTFKAWNLD
    |||||
orf88a      TGMPRDFASDIEVTDKATGEKLERTIRVNHPLTLHGITIIYQASFADGGSDLTFKAWNLD
            270      280      290      300      310      320

30 orf88.pep          100     110     120     130     140     150
    ASREPVLKATSIHQFPLEIGKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLPDVRV
    |||||
orf88a      ASREPVLKATSIHQFPLEIGKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLNDVRV
            330     340     350     360     370     380

35 orf88.pep          160     170     180     190     200     210
    TQEGHKYTNXXXXXXRIRDAPQAVEYKNYMLPVLQEODYFWITGTRSLQQYRWLR
    |||||
orf88a      TQEGHKYTNIGPSIVYRIRDAAGQAVEYKNYMLPVLQEODYFWITGTRSLQQYRWLR
            390     400     410     420     430     440

40 orf88.pep          220     230     240     250     260     270
    PLDKQLKADTFMALREFLKDGEGRKRVADATKGAPAEIREQFMLAAENTLNI FAQKGYL
    |||||
orf88a      PLDKQLKADTFMALREFLKDGEGRKRLVADATKGAPAEIREQFMLAAENTLNI FAQKGYL
            450     460     470     480     490     500

45 orf88.pep          280     290     300     310     320     330
    GLDEFITSNIPKEQQDKMQGYFYEMLYGVMAALDETXYTRYGLPEWQQDEARNRFLHSM
    |||||
orf88a      GLDEFITSNIPKEQQDKMQGYFYEMLYGVMAALDETIRRYGLPEWQQDEARNRFLHSM
            510     520     530     540     550     560

50 orf88.pep          340     350     360     370
    DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTSXXGPLLVL
    |||||
orf88a      DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTSFGALLVYLGSVLLVLGTVLMFYVREKR
            570     580     590     600     610     620

60 orf88a      AWWLFS DGKIRFAMSSARSERDLQKEFPKHVESLQRLGKDLNHDX
            630     640     650     660     670

```

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

```

1   ATGAGTAAAT CCCGTAGATC TCCCCACTT CTTCCCGTC CGTGGTTCGC
51  TTTTTCAGC TCCATGCGCT TTGCGGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT

```

151 TATTTGGTCA AATTCGATC GTTTTGGGCG CAGATTTTGG GTTTTCTGGG  
 201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCTGTTATC ATGATGTTTT  
 251 TGGTGGTTTC TACCAGTTTG TGCCTGATTC GCAATGTGCC GCCGTTCTGG  
 301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC  
 351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA  
 401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA  
 451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG  
 501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTTGCCTG GCGGGTGTGA  
 551 TAGACAGTAA CCTGCTGTTG AACTGGGTA TGCTGACCGG TCGGATGTGT  
 601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT  
 651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC  
 701 AGAGTGCCTG TGTGGTTTTT CTGAATGCCG ACAACGGGAT ATTGGTTCAG  
 751 GACTTGCCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTTCACAA  
 801 TACGGGTATG CCGCGCGATT TTGCCAGTGA TATTGAAGTA ACGGATAAGG  
 851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC  
 901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA  
 951 TTTGACATT C AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTGC  
 1001 TGTTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAAT TGGCAAAACAC  
 1051 AAATATCGTC TTGAGTTCGA TCAGTTTACT TCTATGAATG TGGAGGACAT  
 1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG  
 1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CGGCCCTTCC  
 1201 ATTGTTTACC GTATCCGTGA TCGCGCAGGG CAGGCGGTGC AATATAAAAA  
 1251 CTATATGCTG CCGGTTTTGC AGGAACAGGA TTATTTTGG ATTACCGGCA  
 1301 CGCGCAGCGG CTTGCACGAG CAATACGCTT GGCTGCGTAT CCCCTTGGAC  
 1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAGA  
 1401 TGGGGAAGGG CGCAACGTC TGGTTGCCGA CGCAACCAA GGCACACCTG  
 1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAACAC GCTGAACATC  
 1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAAAT  
 1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT  
 1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG  
 1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT  
 1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC  
 1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTGCAGAT GACCCGTTCC  
 1801 CCGGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC  
 1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTTCA  
 1901 ACGGCAAAAT CCGTTTGGCC ATGTCTTCGG CCCGACGCGA ACGGGATTTG  
 1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA  
 2001 CTTGAATCAT GACTGA

This encodes a protein having amino acid sequence <SEQ ID 332>:

40 1 MSKSRRSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPOTD  
 51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW  
 101 REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVQ GFQGKTINRE  
 151 DGSVLIAAKK GTMNKWGYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV  
 201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ  
 45 251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPHT  
 301 LHGITYQAS FADGGSDLTF KAWNLDASR BPVVLKATSI HQFPLEIGKH  
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS  
 401 IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSGLQQ QYRWLRIPLD  
 451 KQLKADTFMA LREFLKDGEK RKRLVADATK GAPAEIREQF MLAAENTLNI  
 50 501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGVMNAA LDETIRRYGL  
 551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS  
 601 PGALLVYLGS VLLVLGTVLM FYVREKRAWV LFSDBGKIRFA MSSARSERDL  
 651 QKEFPKHVES LQRLGKDLNH D\*

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

55 orf88a.pep MSKSRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60  
 orf88-1 MSKSRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60  
 60 orf88a.pep QIFGFLGLYDVYASAWFVVIIMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120  
 orf88-1 QIFGFLGLYDVYASAWFVVIIMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120  
 orf88a.pep SSLLDVKIAP EVAKRYLEVQGFQGKTINRE DGSVLIAAKKGTMNKWGYIFAHVALIVICL 180  
 65 orf88-1 SSLLDVKIAP EVAKRYLEVQGFQGKTINRE DGSVLIAAKKGTMNKWGYIFAHVALIVICL 180  
 orf88a.pep GGLIDSNLLL KLGMLTGRIVPDNQAVYAKDFKPEILGASNLSFRGNVNISEGQSADVVF 240

-222-

	orf88-1		240
	orf88a.pep	LNADNGILVQDLPEVVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
5	orf88-1		300
	orf88a.pep	LHGITIYQASFADGGSDLTFAKAWNLGDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT	360
10	orf88-1		360
	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
15	orf88-1		420
	orf88a.pep	PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLDGEGRKRLVADATK	480
	orf88-1		480
20	orf88a.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88-1		540
25	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTS	600
	orf88-1		600
30	orf88a.pep	PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88-1		660
	orf88a.pep	LQRLGKDLNHD	672
35	orf88-1		672

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N. gonorrhoeae*:

40	orf88.pep	MVFLNADNGILVQDLPEVVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
	orf88ng	:	60
	orf88.pep	PLTLHGITIYQASFADGGSDLTFAKAWNLGDASREPVLKATSIHQFPLEIGKHKYRLEFD	120
45	orf88ng		120
	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDVRVAVTQEGHKYTNXXXXXXYRIRDAAGQAVEYKN	180
50	orf88ng	:	180
	orf88.pep	YMLPVLQEQQDYFWITGTRSLQQQYRWLRIPLDKQLKADTFMALREFLDGEGRKRLVAD	240
	orf88ng	:    :    :	240
55	orf88.pep	ATKGAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVM	300
	orf88ng		300
60	orf88.pep	NAALDETTRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88ng		360
	orf88.pep	TRSXGPLLVL	371
65	orf88ng	TRSPGALLVYLGSVLLVLGTVMFYVPPKRAWVLFSDGKIRFAMSSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

	1	MVFLNADNGM	LVQDLFFEVK	LKKFHIDFYN	TGMPRDFASD	IEVTDKATGE
5	51	KLERTIRVNH	PLTLHGITIY	QASFADGGSD	LTFKAWNLRD	ASREPVVVLA
	101	TSIHQFPLEI	GKHKYRLEFD	QFTSMNVEDM	SEGAEREKSL	KSTLNDVRAV
	151	TQEGKKYTN	GPSIVYRIRD	AAGQAVEYKN	YMLFILQDKD	YFWLTGTRSG
	201	LQQQYRWLRI	PLDKQLKADT	FMALREFLKD	GEGRKRLVAD	ATKDAPAEIR
	251	EQFMLAAENT	LNIFAQKGYL	GLDEFITSNI	PKGQQDKMQG	YFYEMLYGVM
	301	NAALDETIRR	YGLPEWQQDE	ARNRFLHSM	DAYTGLTEYP	APMLQLDGF
10	351	SEVRSSGLQM	TRSPGALLVY	LGSVLLVLGT	VFMFYVPKKR	AWVLFSNXXI
	401	RFAMSSARSE	RDLQKEFPKH	VESLQRLGKD	LNHD*	

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

	1	ATGAGTAAAT	CCCGTATATC	TCCCACACTT	CTTTCCTCGC	CGTGGTTCGC
15	51	TTTTTTCAGC	TCCATGCGCT	TTGCGGTCGC	TTTGCTCAGT	CTGCTGGGTA
	101	TTGCATCGGT	TATCGGCACG	GTGTTACAGC	AAAACCAGCC	GCAGACGGAT
	151	TATTTGGTCA	AATTCGGACC	GTTTGGACT	CGGATTTTGT	ATTTTTTGGG
	201	TTTGTATGAT	GTCTATGCTT	CGGCATGGTT	TGTCGTATATC	ATGATGTTTC
	251	TGGTGGTTTC	TACCAAGTTG	TGTTTAATCC	GTAACGTTCC	GCCGTTTGG
	301	CGCGAAATGA	AGTCTTTCCG	GGAAAAGGTT	AAAGAAAAAT	CTCTGGCGGC
20	351	GATGCGCCAT	TCTTCGCTGT	TGGATGTAAT	AATTGCCCCC	GAAGTTGCCA
	401	AACGTTATCT	GGAGGTGCGG	GTTTTTCAGG	GAAAAACCGT	CAGCCGTGAG
	451	GACGGTCCG	TTCTGATTGC	CGCCAAAAAA	GGCAaatga	acaaATGGGG
	501	CTATATCTTT	GCccaaqtag	ctTTGATTGT	CATTTGCCTG	GGCGGTTGA
25	551	TAGACACTAA	CCTGCTGCTG	AAGCTGGGTA	TGCTGGCCGG	TCGGATTGTT
	601	CCGGACAATC	AGGCGGTTTA	TGCCAAGGAT	TTCAAGCCCC	AAAGTATTTT
	651	GGGTGCGTCC	AATCTCTCAT	TAGGGGCAA	CGTCAATATT	TCCAGGGGCG
	701	AAAGTGCGGA	TGTGGTTTTC	CTGAATGCCG	ACAACGGGAT	GTTGGTTTCA
	751	GACTTGCTTT	TTGAAGTCAA	ACTGAAAAAA	TTCCATATCG	ATTTTTTACAA
30	801	TACGGGTATG	CCGCGCGATT	TTGCCAGCGA	TATTGAAGTA	ACGGACAAGG
	851	CAACCGGTGA	GAACTTCGAG	CGCACCATCC	GCGTGAACCA	TCCTTTGACC
	901	TTGCACGGCA	TCACGATTTA	TCAGGCGAGT	TTTGCCGACG	GCGGTTCCGA
	951	TTTGACATTC	AAGGCGTGGA	ATTTGAGGGA	TGCTTCGCGC	GAACCTGTCT
35	1001	TGTTGAAGGC	AACCTCCATA	CACCAGTTTC	CGTTGGAAAT	CGGCAAACAC
	1051	AAATATCGTC	TTGAGTTCGA	TCAGTTCACT	TCTATGAATG	TGGAGGCAT
	1101	GAGCGAGGGT	GCGGAACGGG	AAAAAAGCCT	GAAATCCACT	CTGAACGATG
	1151	TCCGCGCCGT	TACTCAGGAA	GGTAAAAAAT	ACACCAATAT	CGGCCCTTCC
	1201	ATCGTGTACC	GCATCCGTGA	TGcggCAGGG	CAGGCGGTCT	AAATATAAAA
40	1251	CTATATGCTG	CCGATTTTGC	AGGACAAAGA	TTATTTTGG	CTGACCGGCA
	1301	CGCGCAGCGG	CTTGACGAG	CAATACCGCT	GGCTGCGTAT	CCCCTTGGAC
	1351	AAGCAGTTGA	AAGCGGACAC	CTTTATGGCA	TTGCGTGAGT	TTTTGAAAGA
	1401	TGGGGAAGGG	CGCAAACGTC	TGTTTGCCGA	CGCAACCAAA	GACGCACCTG
	1451	CGGAAATCCG	CGAACAATT	ATGCTGGCTG	CGGAAACAC	GCTGAATATC
	1501	TTTGCGCAAA	AAGGCTATTT	GGGATTGGAC	GAATTTATTA	CGTCCAATAT
45	1551	CCCGAAAGGG	CAGCAGGATA	AGATGCAGGG	CTATTTCTAC	GAAATGCTTT
	1601	ACGGCGTGAT	GAACGCTGCT	TTGGATGAAA	CCATACGCCG	GTACGGCTTG
	1651	CCCGAATGGC	AGCAGGATGA	AGCGCGGAAC	CGTTTCTGCT	TGCACAGTAT
	1701	GGATGCCTAT	ACGGGGCTGA	CGGAATATCC	CGCGCCTATG	CTGCTCCAGC
	1751	TTGACGGGTT	TTCCGAGGTG	CGTTCTCTAG	GTTTGCAGAT	GACCCGTTCT
50	1801	CCGGGTGCGC	TTTTGGTCTA	TCtcggctcg	gtattgttgg	TTTTGGgtac
	1851	ggtaTttatg	tTTTATGTGC	GCGAAAAACG	GGCGTGGgta	tGTTTTCag
	1901	aCGGCAAAAT	CCGTTTTGCT	ATGtCTTcgg	CCcgcagcga	ACGGGATTTG
	1951	cAGaaggaaT	TTCCAAAACA	CGtcgAGAGC	CTGCAACggc	tcggcaaggA
	2001	CttgaaTCAT	GACTga			

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

55	1	MSKSRIPTL	LSRPWFAFFS	SMRFAVALLS	LLGIASVIGT	VLQQNQPTD
	51	YLVKFGPFWT	RIFDFLGLYD	VYASAEVVI	MMFLVVSLSL	CLIRNVPPFW
	101	REMKSFREKV	KEKSLAAMRH	SSLVDVKIAP	EVAKRYLEVR	GFQGKTVSRE
	151	DGSVLIAAKK	GTMNKWGYIF	AQVALIVICL	GGLIDSNLL	KLGMLAGRIV
	201	PDNQAVYAKD	FKPESILGAS	NLSFRGNVNI	SEGQSADVV	LNADNGMLVQ
60	251	DLFFEVLKLL	FHIDFYNTGM	PRDFASDIEV	TDKATGEKLE	RTIRVNHPLT
	301	LHGITIYQAS	FADGGSDLTF	KAWNLRDASR	EPVVLKATSI	HQFPLEIGHK
	351	KYRLEFDQFT	SMNVEDMSEG	AEREKSLKST	LNDVRAVTQE	GKKYTNIGPS
	401	IVYRIRDAAG	QAVEYKNYML	PIQDKDYFW	LTGTRSGLQ	QYRWLRIPLD
	451	KQLKADTFMA	LREFLKDGE	RKRLVADATK	DAPAEIREQ	MLAAENTLNI
65	501	FAQKGYLGLD	EFITSNIPKG	QQDKMQGYFY	EMLYGVMNAA	LDETIRRYGL

551 PEWQQDEARN RFLLSHSDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS  
 601 PGALLVYLGS VLLVLGTVM FYVREKRAWV LFSDGKIRFA MSSARSERDL  
 651 QKEFPKHVES LQRLGKDLNH D\*

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

```

5      orf88-1.pep  MSKSRRSPPLLSRPWFAFFSSMRFAVALLSLGLIASVIGTVLQONQPQTDYLVKFGSFWA  60
      orf88ng-1    MSKSRIIPTLLSRPWFAPFFSSMRFAVALLSLGLIASVIGTVLQONQPQTDYLVKFGPFWT  60

10     orf88-1.pep  QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH  120
      orf88ng-1    RIFDFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH  120

15     orf88-1.pep  SSLLDVKIAPEVAKRYLEVQGFQGTINREDGSLVIAAKKGTMNKGYIFAQVALIVICL  180
      orf88ng-1    SSLLDVKIAPEVAKRYLEVGRGFQGTVSREDGSLVIAAKKGTMNKGYIFAQVALIVICL  180

20     orf88-1.pep  GGLIDSNLLLKLGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF  240
      orf88ng-1    GGLIDSNLLLKLGLMAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF  240

25     orf88-1.pep  LNADNGILVQDLPEFVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT  300
      orf88ng-1    LNADNGMLVQDLPEFVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT  300

30     orf88-1.pep  LHGITIYQASFADGGSDLTFAKWNLDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT  360
      orf88ng-1    LHGITIYQASFADGGSDLTFAKWNLRDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT  360

35     orf88-1.pep  SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML  420
      orf88ng-1    SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML  420

40     orf88-1.pep  PVLQEQDYFWITGTRSGLQQQYRWIRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK  480
      orf88ng-1    PILQDKDYFWLTGTRSGLQQQYRWIRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK  480

45     orf88-1.pep  GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA  540
      orf88ng-1    DAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVMNAA  540

50     orf88-1.pep  LDETIRRYGLPEWQQDEARNRFLLSHSDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS  600
      orf88ng-1    LDETIRRYGLPEWQQDEARNRFLLSHSDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS  600

55     orf88-1.pep  PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES  660
      orf88ng-1    PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES  660

60     orf88-1.pep  LQRLGKDLNHD  671
      orf88ng-1    LQRLGKDLNHD  671
  
```

Furthermore, ORG88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

```

55     gi|2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537
        Score = 94.4 bits (231), Expect = 2e-18
        Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)

        Query: 16  FAFFSSMRFAVALLSLGLIASVIG-TVLQONQPQTDYLVKFGPFWTRIFDFLGLYDVYAS  74
          + F +S++ A+ ++ +LGI S++G T ++QNQ   YL +FG           L L DV+ S
        Sbjct: 80  YDFLASLKLAFIMLVLGILSMLGSTYIKQNQSFEWYLDQFGYDVGIWIWKLWLNVDVFS  139

60     Query: 75  AWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRHSSLLDVKIAPEVAK  134
          +++++ ++ L V+   C I+ +P W++ S +E++ +   A +H   + VKI P+ K
        Sbjct: 140 WYYILFIVLLAVNLIFCSIKRLPRVWKQAFS-KERILKLDEHAEKHLKPITVKI-PDKDK  197

65     Query: 135 --RYLEVGRGFQGTVSREDGSLVIAAKKGTMNKGYIFAQVALIVICLGLIDSNLLLKL  192
          ++L +GF+   V E   + + A+KG ++ G   +AL+VI G LID
        Sbjct: 198 VLKFLKKKGFK-VFVEEEGNKLYVFAEKGRFSRLGVYITHIALLVIMAGALID-----  249
  
```

Query: 193 GMLAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVFLNADNGMLVQDL 252  
 +I+G RG++ ++EG + DV+ + A+ L  
 Sbjct: 250 -----AIVGV-----RGLSLIVAEGLTNDVMLVGAE--QKPYKL 280

Query: 253 PFEVKLKKFHIDFY---NTGMPRDFA-----SDIEVTDKATGEKLER--TIRVNHPLT 300  
 PF V L F I Y N + + FA SDIE+ + G K+E T++VN P  
 Sbjct: 281 PFAVHLIDFRIKTYAEENPNVDKRFAQAVSSYESDIEIIN---GGKVEAKGTVKVNEPFD 337

Query: 301 LHGITIYQASFA--DGGSDLTFKAWNLRDASREP 332  
 ++QA++ DG S + + + A +P  
 Sbjct: 338 FGRYRLFQATYGILDGTSGMGVIVVDRKKAHEDP 371

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 40

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 337>:

1 ATGATGAGTA ATAmAATGGm ACaaaaAGGG TTTACATTGA TTGmGmTGAT  
 51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT  
 101 ATCmAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG  
 151 GyCGGTATCA ACAATATTTT CAAACAGTTT ATTTTGAAAA ATCCCTTGGA  
 201 CGATAATCAG ACCATCGAGA ACAAACTGGA AATATTTGTC TCAGGCTATA  
 251 AGATGAATCC GAAAATTGCC AAAAaTATA GTGTTTCGGT AAAGTTTGTC  
 301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG  
 351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA  
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA  
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAATAA

This corresponds to the amino acid sequence <SEQ ID 338; ORF89>:

1 MMSNXMXQKG FTLLIXMIVV AILGIISVIA IPSYXSYIEK GYQSOLYTEM  
 51 XGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF  
 101 DKEKSRAAYRL VGVPKAGTGY TSVWMNSVG DGYKCRDAAS AQAHLETLS  
 151 DVGCEAFSNR KK\*

Further work revealed the complete nucleotide sequence <SEQ ID 339>:

1 ATGATGAGTA ATAAATGGA ACaaaaAGGG TTTACATTGA TTGAGATGAT  
 51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT  
 101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG  
 151 GTCGGTATCA ACAATATTTT CAAACAGTTT ATTTTGAAAA ATCCCTTGGA  
 201 CGATAATCAG ACCATCGAGA ACAAACTGGA AATATTTGTC TCAGGCTATA  
 251 AGATGAATCC GAAAATTGCC AAAAATATA GTGTTTCGGT AAAGTTTGTC  
 301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG  
 351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA  
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA  
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAATAA

This corresponds to the amino acid sequence <SEQ ID 340; ORF89-1>:

1 MMSNKMEQKG FTLLIEMIVV AILGIISVIA IPSYQSYIEK GYQSOLYTEM  
 51 VGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF  
 101 DKEKSRAAYRL VGVPKAGTGY TSVWMNSVG DGYKCRDAAS AQAHLETLS  
 151 DVGCEAFSNR KK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with Pile of *N. gonorrhoeae* (accession number Z69260).

ORF89 and Pile protein show 30% aa identity in 120a overlap:

-226-

```

orf89 8  QKGFTLIXXMIVVAILGIISVIAIPSYXSIEKGYQSOLYTEMXGINNISKQFILKNPL- 66
Pile 5  QKGFTLI MIV+AI+GI++ +A+P+Y Y + S+ G + ++ L + +
5 orf89 67 -DDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGYTLVSW 125
Pile 65 PKDNTS-----AGVASSDKIKGKYVQSVTVAKGVVTAEMASTGVNKEIQGKKLSLW 115

```

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

- 10 ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

```

15 orf89.pep 10 20 30 40 50 60
MMSNXMXQKGFTLIXXMIVVAILGIISVIAIPSYXSIEKGYQSOLYTEMXGINNISKQF
orf89a 10 20 30 40 50 60
MMSNKMEQKGFTLIXXXXXXAIXXXXSVIXXXYSYIEKGYQSOLYTEMVGINNISKQX

20 orf89.pep 70 80 90 100 110 120
ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY
orf89a 70 80 90 100 110 120
ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKGTGTGY

25 orf89.pep 130 140 150 160
TLVWVWNSVGDGYKCRDAASAQAHLTLSSDVGCEAFSNRKKX
orf89a 130 140 150 160
TLVWVWNSVGDGYKCRDAASARAHLETLSSDVGCEAFSNRKKX

```

The complete length ORF89a nucleotide sequence <SEQ ID 341> is:

```

30 1 ATGATGAGTA ATAAATGGA AAAAAAGGG TTTACATTGA TTGNGANGNT
51 NATNGNCNTC GCGATACNCN GCNTTANCAG CGTCATTNCN ATNNNTNCNT
101 ATCNNAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACAATATTTC CAAACAGTNT ATTTTGAAAA ATCCCTGGA
201 CGATAATCAG ACCATCAAGA GCAAACTGGA AATATTTGTC TCAGGCTATA
35 251 AGATGAATCC GAAAATTGCC GAAAAATATA ATGTTTCGGT GCATTTTGTC
301 AATGAGGAAA AACCNAGGGC ATACAGCTTG GTCGGCGTTC CAAAGACGGG
351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGC GACGGATACA
401 AATGCCGTGA TGCCGCTTCT GCCCGAGCCC ATTTGGAGAC CTTGTCCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAG

```

- 40 This encodes a protein having amino acid sequence <SEQ ID 342>:

```

1 MMSNKMEQKG FTLIXXXXXX AIXXXSVIX XXXYXSIEK GYQSOLYTEM
51 VGINNISKQX ILKNPLDDNQ TIKSKLEIFV SGYKMNPKIA EKYNVSVHFV
101 NEEKPRAYSL VGVPKGTGTGY TLVWVWNSVG DGYKCRDAAS ARAHLETLSS
151 DVGCEAFSNR KK*

```

- 45 ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

```

50 orf89a.pep 10 20 30 40 50 60
MMSNKMEQKGFTLIXXXXXXAIXXXXSVIXXXYSYIEKGYQSOLYTEMVGINNISKQX
orf89-1 10 20 30 40 50 60
MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSOLYTEMVGINNISKQF

55 orf89a.pep 70 80 90 100 110 120
ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKGTGTGY
orf89-1 70 80 90 100 110 120
ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY

60 orf89a.pep 130 140 150 160
TLVWVWNSVGDGYKCRDAASARAHLETLSSDVGCEAFSNRKKX
orf89-1 130 140 150 160
TLVWVWNSVGDGYKCRDAASAQAHLTLSSDVGCEAFSNRKKX

```



ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N.*

	orf89	MMSNXMXQKGFTLIXXMI VVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNISKQF	60
	orf89ng	:                  :	
	orf89ng	MMSNKMEQKGFTLIEMMI VVTILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF	60
10	orf89	ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVK FVDKEKSRAYRLVGVPKAGTGY	120
	orf89ng	:          :          :	
	orf89ng	ILKNPQDDNDTLKSKLKI FVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY	120
15	orf89	TLSVWMNSVGDGYKCRDAASAQA HLETLSSDVGCEAFS NRKK	162
	orf89ng	:          :	
	orf89ng	TLSVWMNSVGDGYKCRDATSAQAYS DTL SADSGCEAFS NRKK	162

	1	aTGATGAGCA	ATAAAATGGA	ACAAAAAGGG	TTTACATTGA	TTGAGATGAT
20	51	GATAGTTGTC	ACGATACTCG	GCATCATCAC	CGTCATTGCC	ATACCTTCTT
	101	ATCAGAGTTA	TATTGAAAAA	GGCTATCAGT	CCCAGCTTTA	TACGGAGATG
	151	GTCGGTATCA	ACAATGTTCT	CAAAACAGTTT	ATTTTGAAAA	ATCCCCAGGA
	201	CGATAATGAT	ACCTCAAGA	GCAAACTGAA	AATATTGTGC	TACGGCTATA
	251	AGATGAATCC	GAAAAAttgCC	AAAAAATATA	GTGTTTCGGt	aaggtttGTC
25	301	gatGCGGAAA	AACCAAGGGC	ATACAGTTTG	GTCGGCGTTC	CGAACGCGGG
	351	GACGGGTTAT	ACTTTGTCGG	TATGGATGAA	CAGCGTGGCG	GACGGATACA
	401	AATGCCGTGA	TGCCACTTCT	GCCCAGGCCT	ATTCCGACAC	CTTGTCGCGA
	451	GATAGCGGCT	GTGAAGCTTT	CTCTAATCGT	AAAAAATAG	

30

1	MMSNKMEQKG	FTLIEMMIVV	TILGIISVIA	IPSYQSYIEK	GYSQSLYTEM
51	VGINNVLKQF	ILKNPQDDND	TLKSKLKIFV	SGYKMNPKIA	KKYSVSVRFV
101	DAEKPRAYRL	VGVPNAGTGY	TLSVWMNSVG	DGYKCRDATS	AQAYSDTLSA
151	DSGCEAFSNR	KK*			

35 identity in 162 aa overlap:

		10	20	30	40	50	60
	orf89-1.pep	MMSNKMEQKGFTLIEMMIVVAILGIIISVIAIPSYQSYIEKGYQSQLYTEMVGINNISKQF					
	orf89ng	MMSNKMEQKGFTLIEMMIVVTILGIIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF					
40		10	20	30	40	50	60
	orf89-1.pep	ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY					
45	orf89ng	ILKNPQDDNDTLKSKLKIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY					
		70	80	90	100	110	120
	orf89-1.pep	TLSVWMNSVGDGYKCRDAASAQAHALETLSDDVGCEAFSNRKKX					
50	orf89ng	TLSVWMNSVGDGYKCRDATSAQAYSDTLSADSGCEAFSNRKKX					
		130	140	150	160		

Based on this analysis, including the gonococcal motifs and the homology with the known Pile protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that

5 ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 41

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 345>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTGCGCA ACCCTTGGsG CACCG.GTCC GACG.GCAAA
251 AACAAGCGTT GGCCn.AGAA TTCAACCC...
```

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

```

15 1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
    51  RQKAEAYAIP YDFQRM TAL AVGNPWXTXS DXQKQALAXE FQP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
20 51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
    101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
    151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTCGATT TCCAACGTAT
    201 GACCGCATTG GCGGTGCGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
    251 AACAAGCGTT GGCCAAAGAA TTCAAAACCC TGCTGATCCG CACCTATTC
    301 GGCACGATGC TGA AATTAA AAACGCCAAC GTCAACGTCA AAGACAATCC
25 351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
    401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
    451 GGTA AATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
    501 CGTGTACCGC AACCAATTCG GCGAAATTAT CAAAGCGAAA GCGGTGGACG
    551 GACTGATTGC CGAGTTGAAA GCCAAAAACG GCGGCAATA A
```

30 This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

```

1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
51  RQKAEAYAIP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKLKLNAN VNVKDNPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGASLVTVYR NQFGEIIRAK GVDGLIAELK AKNGGK*
```

35 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N.meningitidis*:

```

40      10      20      30      40      50      60
orf91.pep  MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91a     MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
          10      20      30      40      50      60

45      70      80      90
orf91.pep  YDFQRM TALAVGNPWXTXS DXQKQALAXE FQP
          |||||:|||||:|||||:|||||:|||||:|||||
orf91a     YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKLKLNANVNVKDNPIVN
          70      80      90      100     110     120
```

orf91a KGGKEIIVRAEVGVPQGKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK  
130 140 150 160 170 180

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

```

5       1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTGAGCAT
      51 CGGCATGGCA TTGCGCGCCC CTGCCGACGC GGTAAACCAA ATCCGTCAAA
     101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA GCGGTGATGC CACACCGGCC
     151 CGCCAAAAAG CCGAAGCCTA TGGGATTCCC TATTTCGATT TCCAACGTAT
     201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
    10 251 AACAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTAT'TCC
     301 GGCACGATGC TGAATTATAA AAACGCCAAC GTCAACGTCA AAGACAATCC
     351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
     401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
     451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
    15 501 CGTGATACGC AACCAATTTC GCGAAATTAT CAAAGCGAAA GCGGTGGACG
     551 GACTGATTGC CGAGTTGAAG GCTAAAAACG GCAGCAAGTA A
  
```

This encodes a protein having amino acid sequence <SEQ ID 350>:

```

      1 MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLS ILKSGDANTA
     51 RQKAEAYAIP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRITYS
    101 GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPQGKPV NMDFTTYQSG
     151 GKYRTYNVAI EGASLVTYVR NQFGEI IKAK GVDGLIAELK AKNGSK*
  
```

ORF91a and ORF91-1 show 98.0% identity in 196 aa overlap:

```

      10      20      30      40      50      60
    25 orf91a.pep MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91-1 MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
      10      20      30      40      50      60

      70      80      90     100     110     120
    30 orf91a.pep YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRITYSGTMLKLKNANVNVKDNPIVN
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91-1 YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRITYSGTMLKLKNANVNVKDNPIVN
      70      80      90     100     110     120

      130     140     150     160     170     180
    35 orf91a.pep KGGKEIIVRAEVGVPQGKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91-1 KGGKEIIVRAEVGVPQGKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK
      130     140     150     160     170     180

      190
    40 orf91a.pep GVDGLIAELKAKNGSKX
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91-1 GVDGLIAELKAKNGGKX
      190
    45
  
```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*:

```

    50 orf91.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP 60
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91ng VKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP 60

    orf91.pep YFDFQRM TALAVGNPWXTXSDXQKQALAXEFQP 93
    55 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91ng YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRITYSGTMLKFKNATVNVKDNPIVN 120
  
```

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

```

1 VKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
51 RPKAEAYAVP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEI I KAK GIDGLIAELK AKNGGK*

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 353>:

```

1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51 CGGCATGGCA TTTCCTCCC CGGCCGACGC AGTGGGACAA ATCCGCCAAA
101 ACGCCACACA GGTTTTGACC ATCCTCAAAA GCGGCGACGC GGCTTCTGCA
151 CGCCCAAAG CCGAAGCCTA TGCGTTCCC TATTTCGATT TCCAACGTAT
10 201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG TACCGCGTCC GACGCGCAA
251 AACAAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAAATTCAA AAACGCGACC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAGGCGGCA AGGAAATCGT CGTCCGTGCC GAAGTCGGCA
401 TCCCGGTCA GAAGCCGTC AATATGGACT TTACCACCTA CCAAAGCGGC
15 451 GGCAAATACC GTACCTACAA CGTCGCCATC GAAGGCACGA GCCTGGTTAC
501 CGTGTACCGC AACCAATTCG GCGAAATCAT CAAAGCCAAA GGCATCGACG
551 GGCTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAAATA A

```

This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

```

1 MKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
20 51 RPKAEAYAVP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEI I KAK GIDGLIAELK AKNGGK*

```

ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

```

25      10      20      30      40      50      60
orf91-1.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
           |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf91ng-1   MKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP
           10      20      30      40      50      60

30      70      80      90      100     110     120
orf91-1.pep YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLNANVNVKDNPIVN
           |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf91ng-1   YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKFKNATVNVKDNPIVN
           70      80      90      100     110     120

35      130     140     150     160     170     180
orf91-1.pep KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI I KAK
           |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf91ng-1   KGGKEIIVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTVYRNQFGEI I KAK
           130     140     150     160     170     180

40      190
orf91-1.pep GVDGLIAELKAKNGGKX
           |:|||||:|||||
45  orf91ng-1 GIDGLIAELKAKNGGKX
           190

```

In addition, ORF91ng-1 shows homology to a hypothetical *E.coli* protein:

```

50  sp|P45390|YRBC ECOLI HYPOTHETICAL 24.0 KD PROTEIN IN MURA-RPON INTERGENIC
    REGION PRECURSOR (F211) >gi|606130 (U18997) ORF_f211 [Escherichia coli]
    >gi|1789583 (AE000399) hypothetical 24.0 kd protein in murZ-rpoN intergenic
    region [Escherichia coli]Length = 211

    Score = 70.6 bits (170), Expect = 6e-12
    Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)

55  Query: 59 VPYDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKFKNATVNVKDNPI 118
    +PY + AL +G +++A+ AQ++A F+ L + Y + + T + P
    Sbjct: 65 LPYVQVKYAGALVLGQYK SATPAQREAYFAAFREY LKQAYGQALAMYHCQTYQIA--PE 122

60  Query: 119 VNKGGKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQFG 174
    G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S++T +N++G
    Sbjct: 123 QPLGDKTIVPIRVTTIDPNGRPPVRLDFQWRKNSQTGNWQAYDMIAEGVSMITTKQNEWG 182

```

Query: 175 EIIKAKGIDGLIAELKA 191  
 +++ KGIDGL A+LK+  
 Sbjct: 183 TLLRTRKIDGLTAQLKS 199

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 42

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 355>:

```

10      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
     101  TGATCACGCA TACCCTCATC TCAAAATACA GTTTTGnnnn nnnnnnnnnn
     151  nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATT TGGCCGTCAT
     201  CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGCGAA
     251  AAGTCATCGT CTTCGGCAGC CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     301  GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
     451  AACTGATAC AAAAAACCGT AGGCGAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 356; ORF97>:

```

20      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMITHTLI SKYSFGXXXX
      51  XXXXAISKSG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVGE*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 357>:

```

25      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACCCAAAAC GAAACCGCTA
     101  TGACCAACGCA TACCCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC
     151  CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATT TGGCCGTCAT
     201  CGACCATCAG GAAGCCGCCG GCCGAAACGG CTTAACGATG CAGCCGCGAA
     301  AAGTCATCGT CTTCGGCAGC CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     351  GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCCTCG TTACCGAAAC
     401  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     451  GACGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
     451  AACTGATAC AAAAAACCGT AGGCGAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 358; ORF97-1>:

```

      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTTHTLT SKYSFDETVS
      51  RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVGE*
  
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF97 shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) from strain A of *N.meningitidis*:

```

45      orf97.pep      10      20      30      40      50      60
      MKHILPLIAASALCISTASAHASEPSTQNETAMITHTLISKYSFGXXXXXXXXXAISKSG
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf97a      MXHILPLXXASALCISTASXHPASEPQTQNETAMTTHTLTSKYSFDETVSRLETAISKSG
      10      20      30      40      50      60
  
```

-232-

```

              70      80      90      100      110      120
orf97.pep    MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK
              |||||
orf97a       MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK
              70      80      90      100      110      120

              130      140      150      160
orf97.pep    VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIVGEX
              |||||
orf97a       VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
              130      140      150      160

```

The complete length ORF97a nucleotide sequence <SEQ ID 359> is:

```

1  ATGANACACA TACTCCCCCT GANTGNCGCA TCCGCACTCT GCATTTCAAC
51 CGCTTCGNN CATCCTGCCA GCGAACCGCA AACCCAAAAC GAAACCGCTA
15 101 TGACCACGCA TACCCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC
151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATT TGGCCGTCAT
201 CGACCATCAG GAAGCCGCC CCGGAAACGG CTTAACGATG CAGCCGGCAA
251 AAGTCATCGT CTCGGCAGC CCCAAAGCCG GTACGCCGCT GATGGTCAAA
301 GACCCCGCCT TCGCCCTGCA ACTGCCCTG CCGCTCNTCG TTACCGAAAC
20 351 GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
401 GCAGCCGCAT CGGTTTCGAC GAAGTGCGAA ACACTTTGGC AAACGCCGAA
451 AAAGTATAC AAAAACCAT AGGCGAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 360>:

```

1  MXHILPLXXA SALCISTASX HPASEPQTQN ETAMTTHTLT SKYSFDETVS
25 51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
101 DPAFALQLPL RVXVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
151 KLIQKTIGE*

```

ORF97a and ORF97-1 show 95.6% identity in 159 aa overlap:

```

              10      20      30      40      50      60
30 orf97a.pep    MXHILPLXXASALCISTASXHPASEPQTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG
              |||||
orf97-1       MKHILPLIAASALCISTASAHASEPSTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG
              10      20      30      40      50      60

              70      80      90      100      110      120
35 orf97a.pep    MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK
              |||||
orf97-1       MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK
              70      80      90      100      110      120

              130      140      150      160
40 orf97a.pep    VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
              |||||
orf97-1       VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIVGEX
              130      140      150      160
45

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N.*

*gonorrhoeae*:

```

50 orf97.pep    MKHILPLIAASALCISTASAHASEPSTQNETAMITHTLISKYSFGXXXXXXXAIIKSKG 60
              |||||
orf97ng       MKHILPPIAASAFICISTASAHAGKPTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG 60

              70      80      90      100      110      120
55 orf97.pep    MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK 120
              |||||
orf97ng       MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK 120

              130      140      150      160
orf97.pep    VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIVGE 159
              ||:|||||:||||:|||||
60 orf97ng     VRTAYTDTRALIVGSRIISFDEVANTLANAEKLIQKTIVGE 159

```

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>:

```

      1 MKHILPPIAA SAFCISTASA HPAGKPPTQN ETAMTHTLT SKYSFDETVS
5      51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
      101 DPAFALQLPL RVLVTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
      151 KLIQKTVGE*

```

Further work revealed the complete nucleotide sequence <SEQ ID 363>:

```

      1 ATGAAACACA TACTCCCcct gatcgccgca TccgcactCT GCATTCAAC
      51 CGCTTCGGCA CACCTGCGCG GCAAACCGCC CACCCAAAAC GAAACCGCTA
10     101 TGACCACGCA CACCTCACC TCGAAATACA GTTTTGACGA AACCGTCAGC
      151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATT TTGCCGTCAT
      201 CGACCATCAG GAAGCGGCAC GCCGAAACGG CCTGACCATG CAGCCGGCAA
      251 AAGTCATCGT CTTCGGCAGC CCCAAGGCCG GTACGCCgct GATGGTCAAA
      301 GACCCCGCCT TCGCCCTGCA ACTGCCCTG CGCGTCCTCG TTACCGAAAC
15     351 GGACGGCAAA GTACGCACCG CCTATACCGA TACGCGCGCC CTCATCGTCG
      401 GCAGCCGCAT CAGTTTCGAC GAAGTGGCAA AACTTTGGC AAACGCCGAA
      451 AAACGTATAC AAAAAACCGT AGGCGAATAA

```

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>:

```

      1 MKHILPLIAA SALCISTASA HPAGKPPTQN ETAMTHTLT SKYSFDETVS
20     51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
      101 DPAFALQLPL RVLVTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
      151 KLIQKTVGE*

```

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

```

25     orf97-1.pep      10      20      30      40      50      60
      MKHILPLIAASALCISTASAH PASEPSTQNETAMTHTLT SKYSFDETVS RLETAIKSKG
      orf97ng-1        10      20      30      40      50      60
      MKHILPLIAASALCISTASAH PAKKPPTQNETAMTHTLT SKYSFDETVS RLETAIKSKG

30     orf97-1.pep      70      80      90      100     110     120
      MDIFAVIDHQEAARRNGLTM QPAKVIVFGTPKAGTPLMVK DPAFALQLPL RVLVTETDGK
      orf97ng-1        70      80      90      100     110     120
      MDIFAVIDHQEAARRNGLTM QPAKVIVFGTPKAGTPLMVK DPAFALQLPL RVLVTETDGK

35     orf97-1.pep      130     140     150     160
      VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX
      orf97ng-1        130     140     150     160
      VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX
40

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described  
 45 above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein, and that it is a useful immunogen.

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* <SEQ ID 365>:

This corresponds to the amino acid sequence <SEQ ID 366; ORF106>:

Further work revealed the following DNA sequence <SEQ ID 367>:

35 This corresponds to the amino acid sequence <SEQ ID 368; ORF106-1>:

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF106 shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) from strain A of *N. meningitidis*:

```

45          10      20      30      40      50      59
   orf106.pep  MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISRRFQTELPDQ
               ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   orf106a     MAFITRLFKSIKQWLVLPLMLSVLPDAAEGIDVSRAEARIXDGGQLSXXSRFQTELPDQ
               10      20      30      40      50      60

50          60      70      80      90      100     110     119
   orf106.pep  LQQALRRGVPLNFTLSWQLSAPIIASYRFLKGQLIGDDDNIDYKLSFHPLTKRYRVTVGA
               : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```



The complete length ORF106a nucleotide sequence <SEQ ID 369> is:

	1	ATGGCTTTTA	TTACGCGCTT	ATTCAAAAGC	ATTAAACAAT	GGCTTGTGCT
20	51	GCTGCCGATG	CTTTCCGTTT	TGCCGGACGC	GGCGGCGGAG	GGGATAGATG
	101	TGAGCGCGCG	CGAAGCGAGG	ATAACGACG	CTCGGCGAGT	TTCCATNAGN
	151	AGCCGCTTCC	AAACCGAGCT	GCCCGACCAG	GCCGAANNNG	CGNNGNGCCG
	201	GGGCGTGNCG	CTCAACTNTA	CCTTAAAGNTG	GCAGCTTTCC	GCCCCGATAA
	251	TCGCTTCTTA	TCGGTTTNA	TTGGGGCAAC	TGATTGGCGA	TGACGACNAT
25	301	ATTGACTACA	AACTGAGTTT	CCATCCGCTG	ACCAACCGCT	ACCGCGTTAC
	351	CGTCGGCGCG	TTTTCGACAG	ANTACGACAC	CTTGGATGCG	GCGATTGCGG
	401	CGACCGGCGC	GGTTGCCAAC	TGGAAGTCC	TGAACAAAGG	CGCGCTGTCC
	451	GGTGCGGAAG	CAGGGGAAAC	CAAGGCGGAA	ATCCGCTTGA	CGCTGTCCAC
	501	TTCAAAACTG	CCCCAGCCTT	TTCAATCAA	TGCACTTGACT	TCTCAAACAT
	551	GGCAATTGGA	TTCGGTTGG	AAACCTCTAA	ACATCATCGG	GAACAATAA

1	MAFITRLFKS	IKQWLVLLPM	LSVLPDAAAE	GIDVSRAEAR	IXDGGQLSXX
51	SRFQTELPDQ	LQXAXXRGVX	LNXTLKWQLS	APIIASYRFX	LGQLIGDDDX
101	IDYKLSFHPL	TNRYRVTVGA	FSTXYDTLDA	ALRATGAVAN	WKVLNKGALS
151	GAEAGETKAE	IRLTLSSTKL	PKPFQINALT	SONWHLDSGW	KPLNIIGNK*

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N. gonorrhoeae*:

40	orf106.pep	MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISRRFQTELPDQ	59
	orf106ng	MAFITRLFKSIKQWLVLPLILSVLPDAAAEGIAATRAEARITDGGRLSISRRFQTELPDQ	60
45	orf106.pep	LQQALRRGVPLNFTLSWQLSAPIIASYRFLKGQLIGDDDNIDYKLSFHPLTKRYRVTVGA	119
	orf106ng	LQQALRRGVPLNFTLSWQLSAPTIASYRFLKGQLIGDDDNIDYKLSFHPLTNRYRVTVGA	120
50	orf106.pep	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTSTSKLPKPFQINALT	179
	orf106ng	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTSTSKLPKPFQINALT	180
	orf106.pep	SQNWHLDSGWKPLNIIGNK 198	
	orf106ng	SQNWHLDSGWKPLNIIGNK 199	

Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

```

1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
51 GTTGCCGATA CTCTCCGTTT TGCCGGACGC GCGGGCGGAG GGCATTGCCG
101 CGACCCGCGC CGAAGCGAGG ATAACCGACG GCGGGCGGCT TTCCATCAGC
5  151 AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAACAGG CGTTGCGCCG
201 GGGCGTACCG CTCAACTTTA CCTTAAGCTG GCAGCTTTCG GCGCCGACAA
251 TCGCTTCTTA TCGGTTTAAA TTGGGGCAAC TGATTGGCGA TGACGACAAT
301 ATTGACTACA AACTAAGTTT CCATCCGCTG ACCAACCCTG ACCGCGTTAC
351 CGTCGGCGCA TTTTCCACCG ATTACGACAC TTTGGATGCG GCATTGCGCG
10  401 CGACCCGCGC GGTGCGCAAC TGGAAAGTCC TGAACAAAGG CGCGTTGTCC
451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCGCTG CGCTGTCCAC
501 TTCAAACTG CCCAAGCCTT TCCAAATCAA CGCATTGACT TCTCAAAACT
551 GGCATTGGA TTCGGTTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 372>:

```

15 1  MAFITRLFKS IKQWLVLPLI LSVLPDAAAE GIAATRAEAR ITDGGRLSIS
51 SRFQTELPDQ LQQALRRGVP LNFTLSWQLS APTIASYRFK LGQLIGDDDN
101 IDYKLSFHPL TNRYRVTVGA FSTDYDTLDA ALRATGAVAN WKVLNKGALS
151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLDSGW KPLNIIGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 373>:

```

1  ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCCCGCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CCGCGCGGGG GCTgACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
35 201 CACCGCCGAC AAAGACAcCT TGTTCAAAAC CCTGTTCTTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCACT CGACGATGCC gCGCCCGCa TCGGGCTGGT
351 GGTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTGTG
401 GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGCCcAAG
40 451 CTCGCCATCC TGCTGCTG.T GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCCGTCTTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTGCGCCGC GTTCCTGCAC CGGGGG.TGC GCTACGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
45 701 GTTGTTCCTT GAAAAAATAT GCCGGCTGG AACAGCTCGG CGTTTATTCG
751 ATGGGTATTT CGTTCGGCGG GCGCGCATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
851 CCGCTCGCCT CTCGGCAACG GCAGAAATCC CCGCCGCCCT GCTTGCCCTC
901 GCCCTCTGC.TGACCGGCAT TTTCTCGCCC CTTGCCCTCC TCCTGCTGCG
50 951 GGAAACTAC GCCGCCGTCC GGTTTATCGT CGTATCGTGT ATG.TGCCGC

```

-237-

5  
1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTT  
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA  
1101 CCGTCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGGCGAGG CCGCC.GGCG  
1151 CGGCGGTTGC CTGTGCCGCC TCATTCTGGC TGTTTTTTGC CTTCAAGACC  
1201 GAAAGCTCyT GCCGCCTGTG GCAGCCGCTC AAACGCCTGC CGCTTTATCT  
1251 GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA  
1301 CGCCGGCAAA CTATCCCTG TTGCGCGCG TATGGGCGGC ATATCTGGCA  
1351 GGCTGCATCC TGCGCCACCG GAAAGATTG CACAACTGT TTCATTATTT  
1401 GAAAAACAA GGTTTCCCAT TATGA

10 This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV  
51 SVLCLGLDQA YVREYYATAD KDTLFTKTLFL PPLLSAAAIA ALLLSRPSLP  
101 SEILFSLDDA AAGIGLVLEF LSFLPIRFL LVLRMGRAL AFSSAQVLPK  
151 LAILLXLPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR  
15 HAPFSPAVLH RGXRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS  
251 MGISFGGAAL LFQSIFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS  
301 ALCXTGIFSP LASLLLPEY AAVRFIVVSC MXPPLFCTLA EISGIGLNVV  
351 RKTRPIALAT LGALANLLL LGLDRAVPAR PXGAAVACAA SFWLFFAFKT  
401 ESSCRLWQPL KRLPLYLHTL FLTSSAAYT CFGTPANYPL FAGVWAAAYLA  
20 451 GCILRHRKDL HKLFHYLKKQ GFPL\*

Further sequence analysis revealed the complete DNA sequence<SEQ ID 375> to be:

1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC  
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCCGCCG  
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG GCTGACGGTG  
251 TCGGTGTGTG GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC  
151 CACCGCCGAC AAAGACACCT TGTTCAAAC CCTGTTCTCG CCGCCGCTGC  
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTCCCCGCC GTCCCTGCCG  
301 TCTGAAATCC TGTTTTCACT CGACGATGCC GCCGCCGGCA TCGGGCTGGT  
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTGC  
120 401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAG  
451 CTCGCCATCC TGCTGTGCTG CGCGCTGACG GTCGGGCTGC TGCACTTTCC  
501 AGCGAACACC GCCGTCCTGA CCGCGCTTTA CGCGCTGGCA AACCTTGCCG  
551 CCGCCGCTTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCGGG  
601 CACGCACCGT TTTGCCCCGC CGTCCTGCAC CGGGGGCTGC GCTACGGCAT  
35 651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC  
701 GTTGTTCCT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTCG  
751 ATGGGTATTT CGTTCGGCGG GCGGGCATT TGTTCCTCAA GCATCTTTTC  
801 AACGGTCTGG ACACCGTATA TTTCCGCGC AATCGAAGAA AACGCCCCGC  
851 CCGCCCCCCT CTCGGCAACG GCAGAAATCC CCGCCGCCCT GCTTGCTTCC  
40 901 GCCCTCTGCC TGACCGGCAT TTTCTGCCC CTTGCCCTCC TCCTGCTGCC  
951 GGAAAACTAC GCCGCCCTCC GGTTTTATCGT CGTATCGTGT ATGCTGCCGC  
1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTC  
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA  
1101 CCGTCTGCTG CTGGGGCTTG CCGTGCCGTC CGGCGGCGCG CGCGGCGCGG  
45 1151 CGGTTGCCTG TGCCGCCTCA TTCTGGCTGT TTTTGCCTT CAAGACCGAA  
1201 AGCTCCTGCC GCCTGTGGCA GCGCTCAAA CGCTGCCGC TTTATCTGCA  
1251 CACATTGTTC TGCCTGACCT CCTCGGCGGC CTACACCTGC TTCGGCACGC  
1301 CGGCAAACTA TCCCTGTTT GCCGGCGTAT GGGCGGCATA TCTGGCAGGC  
1351 TGCATCTGCG GCCACCGGAA AGATTTCAC AACTGTTC ATTATTGAA  
50 1401 AAAACAAGGT TTCCATTAT GA

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV  
51 SVLCLGLDQA YVREYYATAD KDTLFTKTLFL PPLLSAAAIA ALLLSRPSLP  
101 SEILFSLDDA AAGIGLVLEF LSFLPIRFL LVLRMGRAL AFSSAQVLPK  
55 151 LAILLXLPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR  
201 HAPFSPAVLH RGLRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS  
251 MGISFGGAAL LFQSIFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS  
301 ALCITGIFSP LASLLLPEY AAVRFIVVSC MLPPLFCTLA EISGIGLNVV  
351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFAFKTE  
60 401 SSCRLWQPLK RLPLYLHTLF CLTSSAAYTC FGTPANYPLF AGVWAAAYLAG  
451 CILRHRKDLH KLFHYLKKQG FPL\*

Computer analysis of this amino acid sequence gave the following results:

Homology with EpsM from *Streptococcus thermophilus* (accession number U40830).

- Identities = (25%)

15 Identities = 15/57 (26%), Positives = 31/57 (54%)

Identities = 16/96 (16%), Positives = 36/96 (37%)

*meningitidis:*

35

40 70 80 90 100 110 120

45 orf10a LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLLLPLTVGLLHFPANTAVLTAVYALA

130 140 150 160 170 180

50 |||||:|||| ||||| |||||

orf10a      NLAAAFLLFQNRCLKAVRRAPFSSAVLHRLRYGIPIALSSIAIWGLASADRLFLKKY  
                         190                    200                    210                    220                    230                    240

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

orf10a      AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS  
                250                 260                 270                 280                 290                 300

20

25

30

35

40

45

50

55

60

65

-240-

		70	80	90	100	110	120
	orf10-1.pep	YVREYYATADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE					
	orf10a	YVREYYAAADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE					
5		70	80	90	100	110	120
	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVYALA					
	orf10a	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLXPLTVGLLHFPANTAVLTAVYALA					
10		130	140	150	160	170	180
	orf10-1.pep	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAYWGLASADRLFLKKY					
	orf10a	NLAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAYWGLASADRLFLKKY					
15		190	200	210	220	230	240
	orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
	orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					
20		250	260	270	280	290	300
	orf10-1.pep	ALCXTGIFSPLASLLLPENYAARFIVVSCMXPLFCTLAIEISGIGLNVVRKTRPIALAT					
	orf10a	ALCLTGIFSPLASLLLPENYAARFIVVSCMLPPLFCTLVEISGIGLNVVRKTRPIALAT					
25		310	320	330	340	350	360
	orf10-1.pep	LGALAANLLLLGLDRAVPAR-PXGAAVACAASFVLFVFAKTESSCRLWQPLKRLPLYLHT					
	orf10a	LGALAANLLLLGL--AVPSGGARGAAVACAASFVLFVFAKTESSCRLWQPLKRLPLYMHT					
30		370	380	390	400	410	419
	orf10-1.pep	LFCLTSSAAYTCFGTPANYPLFAGVWAAAYLAGCILRHRKDLHKLFHYLKKQGFPLX					
	orf10a	LFCLASSAAYTCFGTPANYPLFAGVWAVYLAGCILRHRKDLHKLFHYLKKQGFPLX					
35		420	430	440	450	460	470
	orf10-1.pep	MDTKEILGYAAGSIGSAVLAVIILPILLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
	orf10nm	MDTKEILXYAAGSIGSAVLAVIILPILLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
40		420	430	440	450	460	470
	orf10-1.pep	YVREYYAAADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE					
	orf10nm	YVREYYATADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE					
45		70	80	90	100	110	120
	orf10ng.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTSVLTAVYALA					
	orf10nm	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVYALA					
50		130	140	150	160	170	180
	orf10ng.pep	NLAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPIALSSLAYWGLASADRLFLKKY					
	orf10nm	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAYWGLASADRLFLKKY					
55		190	200	210	220	230	240
	orf10ng.pep	AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS					
	orf10nm	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
60		250	260	270	280	290	300
	orf10ng.pep	ALCXTGIFSPLASLLLPENYAARFIVVSCMLPPLFYTLTEISGIGLNVVRKTRPIALAT					
	orf10nm	ALCXTGIFSPLASLLLPENYAARFIVVSCMXPLFCTLAIEISGIGLNVVRKTRPIALAT					
65		310	320	330	340	350	360

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF10 shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) from *N.*

#### 45 *gonorrhoeae*:

	orf10ng.pep	MDTKEILGYAAGSIGSAVLAVIILPILLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
	orf10nm	MDTKEILXYAAGSIGSAVLAVIILPILLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
50	orf10ng.pep	YVREYYAAADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE	120
	orf10nm	YVREYYATADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE	120
55	orf10ng.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTSVLTAVYALA	180
	orf10nm	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVYALA	180
	orf10ng.pep	NLAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPIALSSLAYWGLASADRLFLKKY	240
60	orf10nm	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAYWGLASADRLFLKKY	240
	orf10ng.pep	AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS	300
	orf10nm	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS	300
65	orf10ng.pep	ALCXTGIFSPLASLLLPENYAARFIVVSCMLPPLFYTLTEISGIGLNVVRKTRPIALAT	360
	orf10nm	ALCXTGIFSPLASLLLPENYAARFIVVSCMXPLFCTLAIEISGIGLNVVRKTRPIALAT	360

-241-

```

                    370      380      390      400      410
orf10ng.pep  LGALAAANLLLLLGL--AVPSGGTRGAAVACAASFWLFFVKTESSCRLWQPLKRLPLYMHT
5 orf10nm     LGALAAANLLLLLGLDRAVPAR-PXGAAVACAASFWLFFAFKTESSCRLWQPLKRLPLYLHT
                    370      380      390      400      410

                    420      430      440      450      460      470
10 orf10ng.pep LFCLASSAAYTCFGTPANYPLFAGVWAAAYLAGCILRHRKNLHKLHFHYLKKQGFPLX
orf10nm     LFCLTSSAAYTCFGTPANYPLFAGVWAAAYLAGCILRHRKDLHKLHFHYLKKQGFPLX
                    420      430      440      450      460      470

```

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

```

15 1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCcccgCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG ACTGACGGTG
151 TCGGTATTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CGCCGCCGAC AAAGACACTT TGTTCAAAAC CCTGTTCCCTG CCGCCGCTGC
251 TGTTTTTCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTCGCTG CTGACGATGCC GCGCGCGCA TCGGGCTGGT
351 GGTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTGTG
401 GTATGGAAGG GCGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAA
451 CTCGCCATTC TGCTGCTGTT GCCGCTGACG GTCGGGCTGC TGCACCTTCC
501 GCGGAACACC TCGTCCTGA CCGCGTFTA CCGCTGCGA AACCTTGCCG
25 551 CCGCCGCCTT TTTGCTGTT CAAAACGAT GCCGTCTGAA GGCCGTCGGG
601 CCGCGCCGCT TTTGCTGTT CAAAACGAT GCCGTCTGAA GGCCGTCGGG
651 ACGCTCGCA CTGAGCAGCC TTGCTTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTCTT GAAAAAATAT GCGGGCTGG AACAGCTCGG CGTTTATTCG
751 ATGGGTATTT CGTTCGGCGG GCGGGCATT TGTCTCCAA GCATCTTTTC
30 801 AACGGTCTGG ACACCGTATA TTTTCCGTGC AATCGAAGAA AACGCCACGC
851 CCGCCGCCTT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTTCC
901 GCCCTCTGCC TGACCGGAAT TTTCTGCCC CTCGCCCTCC TCTGTGCTCC
951 GGAAAACTAC GCCGCCGTCC GGTTTACCGT CGTATCGTGT ATGCTGccgc
1001 cgcTGTTTTA CACGCTGACC GAAATCAGCG GCATCGGTTT GAACGTGCTC
35 1051 CGCAAAACGC GTCCGATCGC GCTTGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCGTCTGCTG CTGGGGCTTG CCGTACCGTC CCGCGGCACG CGCGGCGCGG
1151 CGGTGTCCTG TGCCGCCTCA TTCTGGTTGT TTTTGTGTTT CAAGACAGAA
1201 AGCTCCTGCC GCCTGTGGCA GCGCTCAAA CGCCTGCCGC TTTATATGCA
40 1251 CACATTGTTC TGCTGCTGCT CCTCGGCGGC CTACACCTGC TTCGGCACAC
1301 CCGCAAATA CCCcctgtt gccggcgat GGGCGGCATA TCTGGCAGGC
1351 TGCATCTGCG GCCACCGGAA AAATTGTCAC AAAGTGTTC ATTATTTGAA
1401 AAAACAAGGT TTCCCATAT GA

```

This encodes a protein having amino acid sequence <SEQ ID 380>:

```

45 1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYAAAD KDTLFKTLFL PPLLFSAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLEF LSFLPIRFIL LVLRMEGRAL AFSSAQLVFK
151 LAIIIIIIPLT VGLLHFPANT SVLTAVYALA NLAAAIFLLF QNRCRLKAVR
201 RAFPSPAVLH RGLRYGIPLA LSSLAYWGLA SADRLFLKKY AGLEQLGVYS
50 251 MGISFGGAAL LLQSIFSTVW TPYIFRAIEE NATPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFTVWSC MLPPLFYTLT EISGIGLNVV
351 RKTRPIALAT LGALAAANLLL LGLAVPSGGT RGAAVACAAS FWLFFVKTE
401 SSCRLWQPLK RLPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAAAYLAG
451 CILRHRKNLH KLFHYLKKQG FPL*

```

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

```

55 orf10-1.pep 10 20 30 40 50 60
MDTKKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
|||||
orf10ng-1 MDTKKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
10 20 30 40 50 60

60 orf10-1.pep 70 80 90 100 110 120
YVREYYATADKDTLFKTLFLPPLLFSAAIAALLLSRPSLPSEILFSLDDAAAGIGLVLEF
|||||:|||||
orf10ng-1 YVREYYAAADKDTLFKTLFLPPLLFSAAIAALLLSRPSLPSEILFSLDDAAAGIGLVLEF
70 80 90 100 110 120

```

-242-

		130	140	150	160	170	180
5	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLA	ILLPLTVGLLHFPANTAVLTAVYALA				
	orf10ng-1	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLA	ILLPLTVGLLHFPANTSVLTAVYALA				
		130	140	150	160	170	180
10	orf10-1.pep	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGLRYGIPIALSS	IAYWGLASADRFLKKY				
	orf10ng-1	NLAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPIALSS	IAYWGLASADRFLKKY				
		190	200	210	220	230	240
15	orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
	orf10ng-1	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS					
		250	260	270	280	290	300
20	orf10-1.pep	ALCLTGIFSPLASLLLPENYA	AVRFIVVSCMLPPLFCTLA	EISGIGLNVVRKTRPIALAT			
	orf10ng-1	ALCLTGIFSPLASLLLPENYA	AVRFTVVSCMLPPLFYTLTEISGIGLNVVRKTRPIALAT				
		310	320	330	340	350	360
25	orf10-1.pep	LGALAANLLLGLAVPSGGARGA	AAVACAA	FWLFFAFKTESSCRLWQPLKRLPLYLHTLF			
	orf10ng-1	LGALAANLLLGLAVPSGGTRGA	AAVACAA	FWLFFVFKTESSCRLWQPLKRLPLYMHTLF			
		370	380	390	400	410	420
30	orf10-1.pep	CLTSSAAYTCFGTPANYPLFAGV	WAAYLAGCILRHRKDLHKL	FHYLKKQGFP	PLX		
	orf10ng-1	CLASSAAYTCFGTPANYPLFAGV	WAAYLAGCILRHRKNLHKL	FHYLKKQGFP	PLX		
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader peptide and several  
 40 transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6  
 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

45	1..ATCCTGAAAC CGCATAACCA GCTTAAGGAA GACATCCAAC CTGATCCGGC
	51 CGATCAAAAC GCCTTGTC CG AACC GGATGC TCGCAGAGAG GCAGAGCAGT
	101 CGGATGCGGA AAATGCTGCC GACAAGCAGC CCGTTGCCGA TAAAGCCGAC
	151 GAGGTTGAAG AAAAGGCGGG CGAGCCGGAA CGGGAAGAGC CGGACGGACA
	201 GGCAGTGCGT AAGAAAGCGC TGACGGAAGA GCGTGAACAA ACCGTACGGG
50	251 AAAAAGCGCA GAAGAAAGAT GCCGAAACGG TTTAAATACA AGCGGTAAAA
	301 CCGTCTAAAG AAACAGAGAA AAAAGCTTCA AAAGAAGAGA AAAAGGCGGC
	351 GAAGGAAAAA GTTGCACCCA AACCAACCCC GGAACAAATC CTCAACAGCG
	401 GCAgCATCGA AAAGCGCGC AgTGCCGCCG CCAAAGAAGT GCAGAAAATG
	451 AA.AACGTCC GACAAGGCGG AAGC.AACGC ATTATCTGCA AATGGGCGCG
55	501 TATGCCGACC GTCAAGCGCG GGAAGGGCAG CGTGCCAAAC TGGCAATCTT
	551 GGCATATCT TCCAAGGTGG TCGGTATCA GCGGGACAT AAAACGCTTT
	601 ACCGGGTGCA AAGCGGCAAT ATGTCTGCCG ATGCGGTGA

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

60	1..ILKEHNQLKE DIQPD PADQN ALSEPDAATE AEQSDAENAA DKQPVADKAD
	51 EVEEKAGEPE REEPD GQAVR KKALTEEREQ TVREKAQKKD AETVKIQAVK



5	1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGCTGT	CCGGTTTTTT
	51	CTTCGGTTTG	ATACTGGCGA	CGGTCAATTAT	TGCCGGTATT	TGTTTTTATC
	101	TGAACCCAGAG	CGGTCAAAAT	CGCTTCAAAA	TCCCGGCTTC	GTCGAAGCAT
	151	CCTGCGAGAA	CGGAAATCCT	GAAACCGAAA	AACCCGCTTA	AGGAAGACAT
10	201	CCAACCTGAA	CCGGCCGATC	AAAACGCCTT	GTCCGAACCG	GATGCTGCGA
	251	CAGAGGCAGA	GCAGTCGGAT	CGCGAAAAAG	CTGCCGACAA	GCAGCCCGTT
	301	GCCGATAAAG	CCGACGAGGT	TGAAGAAAAG	GCGGCGGAGC	CGGAACGGGA
	351	AGAGCCGGAC	GGACAGGCAG	TGCGTAAGAA	AGCGCTGACG	GAGAGCGGTG
15	401	AACAAACCGT	CAGGGAAAAA	GCGCAGAAGA	AAGATGCCGA	AACGGTTAAA
	451	AAACAAGCGG	TAAAACCGTC	TAAAGAAACA	GAGAAAAAAG	CTTCAAAAGA
	501	AGAGAAAAAG	CGGGCGGAAG	AAAAAGTTGC	ACCCAAACCA	ACCCCGGAAC
	551	AAATCCTCAA	CAGCGGCAGC	ATCGAAAAAG	CGCGCAGTGC	CGCCGCCCAA
20	601	GAAGTCGAGA	AAATGAAAAC	GTCCGACAAG	GCGGAAGCAA	CGCATATATCT
	651	GCAATGGCG	CGGTATGCCG	ACCGTCAGAG	CGGGGAAGGG	CAGCGTGCCA
	701	AACTGGCAAT	CTTGGGCATA	TCTTCCAAGG	TGGTCGGTTA	TCAGGCGGGA
	751	CATAAAACGC	TTTACCGGGT	GCAAAAGCGGC	AATATGTCTG	CCGATCGGGT
25	801	GAATAAAATG	CAGGACGAGT	TGAAAAACA	TGAAGTCGCC	AGCCTGATCC
	851	CTTCTATCGA	AAGCAATAAA			

25	1	MFMNKFSSQS	KGLSGFFFL	ILATVIIAGI	LFYLNQSQGN	AFKIPASSKQ
	51	PAETELPKK	NPKEDIQPE	PADQNALSEP	DAATEAQSD	AEKAADKQPV
	101	ADKADVEEEK	AGPEREED	GQAVRKALT	EEREQTVREK	AQKKDAETVK
	151	KQAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSGS	IEKARSAAAK
	201	EVQKMKTSK	AEATHYLQMG	AYADQSAEG	QRAKLAILGI	SSKVVGYQAG
	251	HKTLYRVOSG	NMSADAVKKM	ODELKKHEVA	SLIRSTESK*	

[illegible]

1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTTTT  
51 CTTCGGTTTG ATACTGGCGA CGGTCATTAT TGCCGGTATT TTGTTTATC

-244-

5  
10  
15

```

101 TGAACCAGAG CGGTCAAAAT GCGTTCAAAA TCCCGGTTCC GTCGAAGCAG
151 CCTGCAGAAA CGGAAATCCT GAAACCGAAA AACCAGCCTA AGGAAGACAT
201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGCTGCGA
251 AAGAGGCAGA GCACTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCCTT
301 GCCGACAAAAG CCGACGAGGT TGAGGAAAAG GCGGACGAGC CGGAGCGGGA
351 AAAGTCGGAC GGACAGGCAG TCGCAAGAA AGCACTGACG GAAGAGCGTG
401 AACAAACCGT CGGGGAAAAA GCGCAGAAGA AAGATGCCGA AACGGTTAAA
451 AAACAAGCGG TAAAACCATC TAAAGAAACA GAGAAAAAAG CTTCAAAGA
501 AGAGAAAAAG GCGGAGAAGG AAAAAGTTGC ACCCAAACCG ACCCGGAAC
551 AAATCCTCAA CAGCGGCAGC ATCGAAAAAG CGCGCAGTGC CGCTGCCAAA
601 GAAGTGCAGA AAATGAAAAC GCCCGACAAG GCGGAAGCAA CGCATTATCT
651 GCAATGGGC GCGTATGCCG ACCGCCGAG CGCGGAAGG CAGCGTGCCA
701 AACTGGCAAT CTTGGGCATA TCTCCAAGG TGGTCGGTTA TCAGGCGGGA
751 CATAAACGC TTTACCGGGT GCAAAGCGGC AATATGTCTG CCGATGCGGT
801 GAAAAAATG CAGGACGAGT TGA AAAACA TGAAGTCGCC AGCCTGATCC
851 GTTCTATCGA AAGCAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 386>:

20

```

1 MFMNKFQSG KGLSGFFFL ILATVIIAGI LFYLNQSGQN AFKIPVPSKQ
51 PAETILKPK NQPKEDIQPE PADQNALSEP DAAKEAEQSD AEKAADKQPV
101 ADKADEVEEK ADEPEREKSD GQAVRKKALT EEREQTVGEK AQKKDAETVK
151 KQAVKPSKET EKKASKEEKK AEKEKVAPKP TPEQILNSGS IEKARSAAAK
201 EVQKMKTPDK AEATHYLQMG AYADRRSAEG QRAKLAILGI SSKVVGQYQAG
251 HKTLRYVQSG NMSADAVKKM QDELKKHEVA SLIRSIESK*

```

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap:

25  
30  
35  
40  
45  
50

```

      10      20      30      40      50      60
orf65a.pep MFMNKFQSGKGLSGFFFLILATVIIAGILFYLNQSGQNAFKIPVPSKQPAETILKPK
           |||||:|||||
orf65-1    MFMNKFQSGKGLSGFFFLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETILKPK
           10      20      30      40      50      60

      70      80      90      100     110     120
orf65a.pep NQPKEDIQPEPADQNALSEPDAAKEAEQSDAEKAADKQPVADKADEVEEKADEPEREKSD
           |||||:|||||
orf65-1    NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDP
           70      80      90      100     110     120

      130     140     150     160     170     180
orf65a.pep GQAVRKKALTEEREQTVGEKAQKKDAETVKKQAVKPSKETEKKASKEEKKAAKEKVAPKP
           |||||:|||||
orf65-1    GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAAKEKVAPKP
           130     140     150     160     170     180

      190     200     210     220     230     240
orf65a.pep TPEQILNSGSIEKARSAAAKEVQKMKTPDKAEATHYLQMGAYADRRSAEGQRAKLAILGI
           |||||:|||||
orf65-1    TPEQILNSGSIEKARSAAAKEVQKMKTSKAEATHYLQMGAYADRQSAEGQRAKLAILGI
           190     200     210     220     230     240

      250     260     270     280     290
orf65a.pep SSKVVGQYQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX
           |||||:|||||
orf65-1    SSKVVGQYQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX
           250     260     270     280     290

```

# 55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N. gonorrhoeae*:

60

```

      30      40      50      60      70      80
ORF65ng IIAGILLYLNQGGQNAFKIPAPSKQPAETILKLNQPKEDIQPEPADQNALSEPDAKE
           |||:|||||:|||||:|||||:|
ORF65    ILKPHNQLKEDIQPPADQNALSEPDAATE
           10      20      30

```

5

ORF65ng AEQSDAEKAAADKQPVADKADVEEEKAGEPEREEDPGQAVRKKALTEEREQTVREKAQKKD  
|||::|:|||||

ORF65 AEQSDAENAADKQPVADKADVEEEKAGEPEREEDPGQAVRKKALTEEREQTVREKAQKKD  
40 50 60 70 80 90

10

ORF65ng AETVKKKAVKPSKETTEKKASKEEKKAAKEKVAPKPTPEQILNSRSIEKARSAAAKEVQKM  
|||:|:|||||

ORF65 AETVKIQAVKPSKETTEKKASKEEKKAAKEKVAPKPTPEQILNSGSIEKARSAAAKEVQKM  
100 110 120 130 140 150

15

ORF65ng KNFGQGGSQRIICKWARMNPNGARKGSVPNWQSWAYLPKWSAIRRDIKRFTACKAAICPP  
| |||:|:|||||

ORF65 XNVRQGGSXRIICKWARMNPVRARKGSVPNWQSWAYLPRWSVIRRDIKRFTGCKAAICLP  
160 170 180 190 200 210

20

ORF65ng MR  
||

ORF65 MR

25	1	MF MNKFQS QS	KGLSGFFFL	<u>ILATV I IAGI</u>	LLYLNQGGQN	AFKIPAPSKQ
	51	PAETELILKL	NQPKEDIQPE	PADQNALSEP	DVAKEAEQSD	AEKAADQPV
	101	ADKADVEEEK	AGEPEREEDP	GQAVRKKALT	EEREQTVREK	AQKKDAETVQ
	151	KKAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
	201	EVQKMNFGQ	GGSRIICKW	ARMPNPGARK	GSVPNWQSWA	YLPKWSAIRR
	251	DIKRETFACKA	AICPPMR*			

	1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGTCTGT	CCGTTTCTCT
	51	CTTCGTTTTG	ATACTGGCAA	CGGTCATTAT	TGCCCGTATT	TTGCTTTATC
35	101	TGAACAGGG	CGGTCAAAAT	CGCTTCAAAA	TCCCGGCTCC	GTCGAAGCAC
	151	CCTGACGAAA	CGGAAATCCT	GAAACTGAAA	ATCCAGCCTA	AGGAAGACAT
	201	CCAACCTGAA	CCGGCCGATC	AAAACGCCTT	GTCCGAACCG	GATGTTGCGA
	251	AAGAGGCAGA	CGAGTCGGAT	CGCGAAAAAG	CTCGGCACAA	GCAGCCCGTT
	301	GCCGACAAag	cgcagcAGGT	TGAAGAAAag	cGGGcgAgc	cggaaCCGGga
	351	aGAGCCGGAC	ggACAGGCAG	TGCGCAAGAA	AGCACTGAcg	gAAGAgcGTG
40	401	AACAAACcgt	cagggAAAAA	GCGCagaaga	AAGATGCCGA	AACGgTTAAA
	451	AAacaaGCgg	tAaaaccgctc	tAAAGAAACa	gagaaaaaaag	cTtcaaaaga
	501	agagaaaaag	gcggcgaaaag	aaaAAGttgc	accctaaaccg	accctggaaAC
	551	aaatcctcaa	cagccgCagc	atcgaaaaaag	cgcgtagtgC	cgctgccaaa
	601	gaAgtgcaGA	AAatgaaaaa	ctTtgggcaa	ggcgGaaGCC	aacgcataAT
	651	CTGcaaatgg	gcgcgtatgc	cgaccgtccg	gagcgcgggaA	gggcagcgtg
45	701	ccaaACtggc	aAtcttGgc	atatctTccg	aagtgctcgG	CTATCAGGCG
	751	GGACATAAAA	CGCTTTACCG	CGTGCAAagc	GGCAatatgt	ccgccgatgc
	801	gGTGAAAAAA	ATGCAGGACG	AGTTGAAAAA	GCATGGGGtt	gcCAGCCTGA
	851	TCCGTGcgAT	TGAAGGCAAA	TAA		

50	1	MFMNKFQSQG	KGLSGFFFL	<u>ILATVIIAGI</u>	LLYLNQGGQN	AFKIPAPSKQ
	51	PAETELILKK	NQPKEDIQPE	PADQNALSEP	DVAKEAQSD	AEKAAADQPV
	101	ADKADVEEKK	AGEPEREEP	GQAVRKALT	EEREQTVREK	AQKKDAETVP
	151	KQAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
55	201	EVQKMKNFGQ	GGSQRIICKW	ARMPTVRSAE	GQRAKLAILEG	ISSEVVGYYA
	251	GHKTLRYRVQ	GNSADAVKK	MQDELKKHGV	ASLRATEGIE	*

60

orf65-1.pep	MFMNKF	10	20	30	40	50	60
	QSGKLSGFFGLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETELKPK						
orf65ng-1	MFMNKF	10	20	30	40	50	60
	QSGKLSGFFGLILATVIIAGILLYLNQGGQNAFKIPAPSKQPAETELKLK						

-246-

		70	80	90	100	110	120
	orf65-1.pep	NQPKEDIQPEPADQNALSEPDAAEAEQSDAEKAADKQPVADKADVEVEEKAGEPEREEDP					
5	orf65ng-1	NQPKEDIQPEPADQNALSEPDVAKEAEQSDAEKAADKQPVADKADVEVEEKAGEPEREEDP					
		70	80	90	100	110	120
	orf65-1.pep	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAKEKVAPKP					
10	orf65ng-1	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAKEKVAPKP					
		130	140	150	160	170	180
	orf65-1.pep	TPEQILNSGSIEKARSAAAKEVQKMKTSKDAEATHYL-QMGAYADRQSAEQRAKLAAILG					
15	orf65ng-1	TPEQILNSRSIEKARSAAAKEVQKMKNFQGGSQRIICKWARMPTVRSAEQRAKLAAILG					
		190	200	210	220	230	240
	orf65-1.pep	ISSKVVGyQAGHKTLyRVQSGNMSADAVKKMQDELKKHVASLIRSIESKX					
20	orf65ng-1	ISSEVVGyQAGHKTLyRVQSGNMSADAVKKMQDELKKHGVASLIRAIEGKX					
		240	250	260	270	280	290

- 25 On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 391>:

	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTACTCG	GTkTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GcGTTTGs.s
	101	TCCAACCTCCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
35	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTCTTT	CCTTGCGCGC	AAAAATCGAG	AAaATCGGCA	AACCGATATG
	351	CGGGAACCTG	AACCCGATAC	TCAACCGGCT	GTTACCCATA	AAATCCATAC
40	401	CCGCCTGCCT	tGCGgTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
	451	GTTTACAGCG	CGTCGCTTTA	CGCGCTGGGA	AgCGGTAGTG	CGGCAACGGG
	501	CGGGTTATAT	ATGCTTGCCCT	TTGCACTGGG	TACGCTGCC	AATCTTtTAG
	551	CAATCGGCAT	TTTtTCCCTG	CAACTGAaAw	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
	651	TGCCGTCCTG	TGGCTGTAA			

- 45 This corresponds to the amino acid sequence <SEQ ID 392; ORF103>:

	1	MNHDITFLTL	FLLGXFGGTH	CIGMCGGLSS	AFXXQLPPhi	NRFWLILLLN
	51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS
	101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPI	KSIPACLAVG	ILWGWLPcGL
50	151	VYSASLYALG	SGSAATGGLY	MLAFALGTLP	NLLAIGIFSL	QLXKIMQNRy
	201	IRLCTGLSVS	LWALWKLAVL	WL*		

Further work elaborated the DNA sequence <SEQ ID 393> as:

	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTACTCG	GTTTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
55	101	TCCAACCTCCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTCTTT	CCTTGCGCGC	AAAAATCGAG	AAAAATCGGCA	AACCGATATG

-247-

5  
351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC  
401 CCGCCTGCCT TGCAGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG  
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CCGCAACGGG  
501 CGGGTTATAT ATGCTTGCTT TTGCACTGGG TACGCTGCCC AATCTTTTAG  
551 CAATCGGCAT TTTTCCCTG CAACTGAAAA AAATCATGCA AAACCGATAT  
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT  
651 TGCCGTCCTG TGGCTGTAA

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

10  
1 MNHDITFTL FLLGFFGGTH CIGMCGGLSS AFALQLPPI NREWLILLN  
51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVL QNILYTAANL LLLFLGLYLS  
101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLA VG ILWGWLP CGL  
151 VYSASLYALG SGSAATGGLY MLAFALGTL NLLAIGIFSL QLKIMQNR  
201 IRLCTGLSVS LWALWKLAVL WL\*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
20	orf103.pep	MNHDITFTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPI	NREWLILLN	TGRVSSYTAI
	orf103a	MNXDITFTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPI	NRXWLILLN	TGRVSSYTAI
		10	20	30	40	50	60
25	orf103.pep	70	80	90	100	110	120
		GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS	SGISSLAAKIE	KIGKPIWRNL
	orf103a	70	80	90	100	110	120
		GLILGLIGQV	GVSLDQTRVX	QNILYTAANL	LLLFLGLYLS	SGISSLAAKIE	KIGKPIWRNL
		70	80	90	100	110	120
30	orf103.pep	130	140	150	160	170	180
		NPILNRLPI	KSIPACLA VG	ILWGWLP	CGLVYSASLY	ALGSGSAAT	GGLYMLAFALGTL
	orf103a	130	140	150	160	170	180
		NPILNRLPI	KSIPACLA VG	ILWGWLP	CGLVYSASLY	ALGSGSAAT	GGLYMLAFALGTL
		130	140	150	160	170	180
35	orf103.pep	190	200	210	220		
		NLLAIGIFSL	QLKIMQNR	YIRLCTGLSV	SLWALWKLAVL	WLX	
40	orf103a	190	200	210	220		
		NLXAIGIFSL	QLKIMQNR	YIRLCTGLSV	SLWALWKLAVL	WLX	
		190	200	210	220		

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

45  
1 ATGAACCANG ACATCACTTT CCTCACCTG TTCCTACTCG GTTCTTTCGG  
51 CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GCGTTTGC GC  
101 TCCAACCTCC CCCGCATATC AACCGCTTNT GGCTGATCCT GCTGCTTAAC  
151 ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCGTGATAC TCGGATTAAT  
201 CGGACAGGTC GGCCTTTTAC TCGACCAAAC CCGCGTCNTG CAGAATATTT  
251 TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC  
301 GGTATTTCTT CTTGGCGGC AAAAATCGAG AAAATCGGCA AACCGATATG  
351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC  
401 CCGCCTGCCT TGCAGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTA  
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CCGCAACGGG  
501 CGGGTTATAT ATGCTTGCTT TTGCACTGGG TACGCTGCCC AATCTTTNGG  
551 CAATCGGCAT TTTTCCCTG CAACTGNAAA AAATCATGCA AAACCGATAT  
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT  
55  
651 TGCCGTCCTG TGGCTGTAA

This encodes a protein having amino acid sequence <SEQ ID 396>:

1 MNXDITFTL FLLGFFGGTH CIGMCGGLSS AFALQLPPI NRXLILLN  
51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVX QNILYTAANL LLLFLGLYLS  
101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLA VG ILWGWLP CGL

-248-

151 VYSASLYALG SGSAAATGGLY MIAFALGTLP NLXAIGIFSL QLXKIMQNRV  
 201 IRLCTGLSVS LWALWKLAVL WL\*

ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

		10	20	30	40	50	60
5	orf103a.pep	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLILLNTGRVSSYTAI					
	orf103-1	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI					
		10	20	30	40	50	60
10	orf103a.pep	GLILGLIGQVGVSLDQTRVXQNILYTAANLILLFLGLYLSGISSLAAKIEKIGKPIWRNL					
	orf103-1	GLILGLIGQVGVSLDQTRVLQNILYTAANLILLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
15	orf103a.pep	NPILNRLLPIKSIPACLA VGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
	orf103-1	NPILNRLLPIKSIPACLA VGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
		130	140	150	160	170	180
20	orf103a.pep	NLXAIGIFSLQLXKIMQNRVIRLCTGLSVSLWALWKLAVLWLX					
	orf103-1	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		
25	orf103a.pep	NLXAIGIFSLQLXKIMQNRVIRLCTGLSVSLWALWKLAVLWLX					
	orf103-1	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N.*

30 *gonorrhoeae*:

	orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFWLILLNTGRVSSYTAI	60
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI	60
35	orf103.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLILLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTAANLILLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
40	orf103.pep	NPILNRLLPIKSIPACLA VGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP	180
	orf103ng	NPILNRLLPIKSIPACLA VGILWGWLPCGLVYSASLYALGSGSATGGLYMLAFALGTLP	180
	orf103.pep	NLLAIGIFSLQLXKIMQNRVIRLCTGLSVSLWALWKLAVLWL	222
45	orf103ng	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWL	222

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTGCTCG	GTTTCTTCGG
	51	CGGAACCTCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
50	101	TCCAACCTCCC	CCCGCATATC	AACCGCTTTT	GGCTGATTCT	GCTGCTTAAC
	151	ACAGGACCGGA	TAAGCAGCTA	TACGGCAATC	GGCCTGATGC	TCGGATTAAT
	201	CGGACAACCTC	GGCATTTCAC	TCGACCAAAc	ccgcgTCCTG	CAAAATATTT
	251	tatacacagc	ctccaaCCTC	CTGCTGCTCT	TTTtaggCCT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG
	351	GCGCAACCTG	AACCGGATAC	TCAACCGGCT	GCTGCCCATC	AAATCCATAC
55	401	CCGCCTGCCT	TGCTGTGCGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
	451	GTTTACAGCG	CATCACTTTA	CGCGCTGGGA	AGCGGTAGTG	CGACAACCGG
	501	CGGACTGTAT	ATGCTTGCCT	TTGCACTGGG	TACGCTGCCC	AATCTTTTGG
	551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACAGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAGCT
60	651	TGCCGTCCTG	TGGCTGTAA			

This encodes a protein having amino acid sequence <SEQ ID 398>:

1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPhi NREWLILLN  
 51 TGRISSTAI GLMLGLIGQL GISLDQTRVL QNILYTASNL LLLFLGLYLS  
 101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLAVG ILWGWLPCGL  
 151 VYSASLYALG SGSATTGGLY MAFALGTLP NLLAIGIFSL QLKIMQNR  
 201 IRLCTGLSVS LWALWKLAVL WL\*

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

		10	20	30	40	50	60
	orf103-1.pep	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhiNREWLILLNLTGRVSSYTAI					
10	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhiNREWLILLNLTGRISSTAI					
		10	20	30	40	50	60
	orf103-1.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
15	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTASNLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
	orf103-1.pep	NPILNRLPIKSIPACLAVGILWGWLPCGLVYSASLYALGSGSAATGGLYMAFALGTLP					
20	orf103ng	NPILNRLPIKSIPACLAVGILWGWLPCGLVYSASLYALGSGSATTGGLYMAFALGTLP					
		130	140	150	160	170	180
	orf103-1.pep	NLLAIGIFSLQLKKIMQNRIRLCTGLSVSLWALWKLAVLWLX					
25	orf103ng	NLLAIGIFSLQLKKIMQNRIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 47

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

1 ATGGAACACC AAAGGCCGCT CCTAGGCTTT CGCTTGGCAC TTTTGGCGGC  
 51 GATGACGTGG GGAACGCTGC CGAT.TCCGT GCGGCAGGTA TTGAAGTTTG  
 101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGCG GCGCGCGGTA  
 151 TTGTTTGTGTT TGCTGGCACT GGGCGGGCGG CTGCcGAAGC GGCGaGGATT  
 201 TTTCTTGGTG CTCATTGAGG CTGCTGCTGC TCGGCGTGCG GGGCATTTCG  
 251 GCAAACCTTG TGCTGATTGC CCAAGGGCTG CATTATATTT CGCCGACCAC  
 301 GACGCAGGTT TTGTGGCAGA TTTCGCGGTT TACGATGATT GTwGTCCGTTG  
 351 TGTGGGTGTT TAAAGACCGG ATGACTGCCG CTCAGAAAAT CGGCTTGCTT  
 401 TTGCTGCTTG CCGGTTTGCT TATGTATTTT AACGATAAAT TCGGCGAGTT  
 451 GTCGGGTTTG GCGCGGTATG C.AAGGGCGT GTTGCTGTGT GCGGCAGGCA  
 501 GTATGGCATG GGTGTGTAAT GCCGTGGCGC AAAAGCTGCT GTCGGCGCAA  
 551 TTCGGGGCGC AACAGATTCT GCTGTTGATT TATGCGGCAA GTGCCGCGCT  
 601 GTTCCTGCCG TTTGCCGAAC CGGCACACAT CGGAAGTATG GACGGTACGT  
 651 TGGCGTGGGT ATGTATTGCG TATTGCTGCT TGAATACGTT AATCGGTTAC  
 701 GGCTCGTTTCG GCGAGGCGTT GAAACATTGG GAGGCTTCCA AAGTCAGCGC  
 751 GGTAAACAACC TTGCTCCCGG TGTTTACCGT AATAAATACT TTGCTCGGGC  
 801 ATTATGTGAT GCCTGAAACT TTTGCCGCGC CGGA..

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

1 MENQRPLLGF RLALLAAMTW GTLPXSVRQV LKFVDAPTLV WVRFTVAAAV  
 51 LFVLLALGGR LPKRRDFSWC SFRLLLLQVA GISANFVLIA QGLHYISPTT  
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLGLL MYFNDKFGEL  
 151 SGLGAYXKGV LLCAAGSMAW VCNAVAQKLL SAQFGPQQIL LLIYAASAAV  
 201 FLPFAEPAHI GSMDGTLAWV CIAYCCLNTL IGYGSFGEAL KHWEASKVSA

251 VTLLPVFTV INTLLGHYVM PETFAAP...

Further work revealed further partial DNA sequence <SEQ ID 401>:

```

5      1 ATGGAACACC AAAGGCCGCT CCTAGGCTTC GCGTTGGCAC TTTTGGCGGC
      51 GATGACGTGG GGAACGCTGC CGATTGCCGT GCGGCAGGTA TTGAAGTTTG
10    101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GGCGGCGGTA
      151 TTGTTTGTTC TGCTGGCACT GGGCGGGCGG CTGCCGAAGC GGCGGGATTT
      201 TTCTTGCTGC TCATTACAGC TGCTGCTGCT CGGCGTGGCG GGCATTTTCGG
      251 CAAACTTTGT GCTGATTGCC CAAGGGCTGC ATTATATTTC GCCGACCACG
      301 ACGCAGGTTT TGTGGCAGAT TTCGCCGTTT ACGATGATTG TTGTCGGTGT
15    351 GTTGGTGTTC AAAGACCGGA TGACTGCCGC TCAGAAAATC GGCTTGGTTT
      401 TGCTGCTGTC CGGTTTGCTT ATGTTTTTTA ACGATAAATT CGGCGAGTTG
      451 TCGGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CCGCAGGCAG
      501 TATGGCATGG GTGTGTTATG CCGTGGCGCA AAAGCTGCTG TCGGCGCAAT
      551 TCGGGCCGCA ACAGATTCTG CTGTTGATTT ATGCGGCAAG TGCCGCCGTG
20    601 TTCTGCCGTG TTGCCGAACC GGCACACATC GGAAGTTTGG ACGGTACGTT
      651 GCGGTGGGTT TGTTTTGCGT ATTGCTGCTT GAATACGTTA ATCGGTTACG
      701 GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG
      751 GTAACAACCT TGCTCCCGT GTTTACCGTA ATAwTwWCTT TGCTCGGGCA
      801 TTATGTGATG CTGAAACTT TTGCCGCGCC GGA...

```

20 This corresponds to the amino acid sequence <SEQ ID 402; ORF104-1>:

```

25    1 MENQRPLLGF ALALLAAMTW GTLPiAVRQV LKFVDAPTLV WVRFTVAAAV
      51 LFLVLLALGGR LPKRRDFSWC SFRLLLLLVA GISANFVLI QGLHYISPTT
      101 TQVLWQISPF TMIVVGVLFV KDRMTAAQKI GLVLLLAGLL MFENDKFGEL
      151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQOIL LLIYAASAAV
25    201 FLFFAEPAHI GSLDGTIAWV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA
      251 VTLLPVFTV IXLLGHYVM PETFAAP...

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

ORF104 and HI0878 show 40% aa identity in 277aa overlap:

```

30    orf104  4  QRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 62
      Q+PLLGF AL+ AM WG+LP +++QVL ++A T+VW P
      HI0878  3  QQPLLGFTFALITAMAWGSLPALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

      orf104  63  --KRRDFSWCSFRLLLLLVAGISANFVLIQGLHYISPTTTQVLWQISPFMTIVVGVLFV 120
      K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
35    HI0878  63  LMKVRQYAW----IMLIGVIGLTSNFFLFSSSLNYIEPSVAQIFIHLSFGLICGVLF 118

      orf104  121 KDRMTAAQKIXXXXXXXXXXXMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL 180
      K+++ QKI ++FND+F +GL Y GV+L G++ WV +AQKL+
40    HI0878  119 KEKLGHLQKIGLFLLLIGLGLFNDREDAFAGLNQYSTGVILGVGGALIWWAYGMAQKLM 178

      orf104  181 SAQFGPQOILLIYAASAAVFLFFAEPAHIGSMDGTIAWVCIAVCCLNTLIGYGSFGEAL 240
      +F QQILL++Y A F+P A+ + + LA +C YCCLNTLIGYGS+ EAL
45    HI0878  179 LRKFNSQQIILLMYLGCIAIFMPMADFQVQELT-PLALICFIYCCCLNTLIGYGSYAEAL 237

      orf104  241 KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277
      W+ SKVS V TL+P+FT++ + + HY P FAAP
      HI0878  238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAP 274

```

50 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N.*

*meningitidis*:

```

55    orf104.pep  10      20      30      40      50      60
      MENQRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWVRETVAAGVLFVLLALGGR
      ||||| : |||||
      orf104a    10      20      30      40      50      60
      MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRETVAAGVLFVLLALGGR
      70      80      90     100     110     120

```



	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPTMTIVVGVLVF
	orf104a	LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPTMTIVVGVLVF
5		70 80 90 100 110 120
	orf104.pep	KDRMTAAQKIGLVLLAGLLMYFNDFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL
	orf104a	KDRMTAAQKIGLVLLAGLLMFFNDFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
10		130 140 150 160 170 180
	orf104.pep	SAQFGPQQIILLIYAASA AVFLPFAEPAHIGSMDGTLAWVCIAYCCINTLIGYGSFG EAL
	orf104a	SAQFGPQQIILLIYAASA AVFLPFAELAHIGSLDGT LAWVCFAYCCINTLIGYGSFG EAL
15		190 200 210 220 230 240
	orf104.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP
	orf104a	KHWEASKVSAVTTLLPVFTVIFSLGHYVMPDPTFAAPDMNGLGYAGALVVVGGAVTA AVG
20		250 260 270 280 290 300

25	1	ATGGA AAAACC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTTGGCGGC
	51	GATGACGTGG	GGAACGCTGC	CGATTGCGGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATCGCGC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGCC	GGCGGCGGTA
	151	TTGTTTGTTT	TGCTGGCATT	GGGCGGGCGG	CTCCGGAAGT	GCGCGGATTT
30	201	TTCTTGCTGC	TCATT CAGGC	TGCTGCTGCT	CGGCGTGGCG	GGCATTTCGG
	251	CAAAC TTTGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTTC	GCCGACCACG
	301	ACG CAGGTTT	TGTGGCAGAT	TTCCGCCGTT	ACGATGATTG	TTGTCCGGTT
	351	GTTGGTCTTT	AAAGACCGGA	TGACTGCCGC	TCAGAAAATC	GGCTTGCTTT
35	401	TGCTGCTTGC	CGGTTTGCTT	ATGTTTTTTA	ACGATAAAAT	CGGCGAGTTG
	451	TCGGGTTTGG	GCGCGTATCG	GAAGGGCGTG	TTGCTGTGTG	CGGCAGGCAG
	501	TATGGCATGG	TCAGTATTATG	CCGTGGCGCA	AAAGCTGCTG	TCGGCCCAAT
	551	TCGGGCCGCA	ACGATTTCTG	CTGTTGATTT	ATTCCGGCAAG	TCGCCGCGTG
40	601	TTCTTGCGCT	TTGCCGA ACT	GGCACACATC	GGAAGTTTGG	ACGGTACGTT
	651	GGCGTGGGTT	TGTTTTCGCT	ATTGCTGCTT	GAATACGTTA	ATCGGTTTACG
	701	GCTCGTTCCG	CAGGCGTTTG	AAACATTTGG	AGGCTTCCAA	AGTCAGCGCG
	751	GTAACAACCT	TGCTCCCCGT	GTTTACCGTA	ATATTTTCTT	TGCTCGGGCA
	801	TTATGTGATG	CCTGATACTT	TTGCCGCGCC	GGATATGAAC	GGTTTGGGTT
	851	ATGCGGCGGC	ACTGGTCTGT	GTCGGGGGTG	CGGTTACGGC	GGCGGTGGGG
	901	GACAGGCTGT	TCAAACGCCG	CTAG		

45 1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTIV WVRFTVAAAV  
51 LFVLLALGGR LPKWRDFSWC SFRLLLLGVA GISANFVLIA QGLHYISPTT  
101 TQVLLQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFNDKFGEL  
151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV  
201 FLFPAELAHV GSLDGTLAWV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA  
251 VTTLPLPVFTI IFSLGLGHYVM PDTFAAPDMN GLGYAGALV VGGAVTAAVG  
50 301 DRLEFKRR\*

		10	20	30	40	50	60
	orf104a.pep	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR					
55	orf104-1						
		MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR					
		10	20	30	40	50	60
60	orf104a.pep	LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF					
	orf104-1						
		LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF					
		70	80	90	100	110	120
65	orf104a.pep	KDRMTAAQKIGLVLVLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAOKLL					

[illegible]

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

20	orf104.pep	MENQRPLLGFRLALLAAMTWGTLPSXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR         :	60
	orf104ng	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
25	orf104.pep	LPKRRDFSWCSEFRLLLLGVAGISANFVLIAQGLHYISPSTTTQVLWQISPFTMIVGVLVF         :	120
	orf104ng	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPSTTTQVLWQISPFTMIVGVLVF	120
30	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL         :	180
	orf104ng	KDRMTAAQKIGLVLLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180
35	orf104.pep	SAQFGPQQIILLIYAASA AVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL         :	240
	orf104ng	SAQFGPQQIILLIYAASA AVFLXAEPAHIGSLDGT LAWVCFVYCCCLNTLIGYGSFGEAL	240
	orf104.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP         :	277
	orf104ng	KHWEASKVSAVTTLLPVFTVIFSLGHYVMPDTFAAPDMNGLGYGALVVVGGAVTA AVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEQ ID 406>:

45

1	<u>MENORPLLGF</u>	<u>ALALLAAMTW</u>	GTLPIAVRQV	LKFVDAPTIV	WVRFTVAAAV
51	LFVLLALGGR	LPKRDRFSWH	SERLLLLGVT	GISANFVLIA	QQLHYISPTT
101	<u>TQVLWQISPF</u>	<u>TMIVVGVLVF</u>	<u>KDRMTAAQKI</u>	<u>GLVLLLVGLL</u>	<u>MFNDHKFGEAL</u>
151	SGLGAYAKGV	LLCAAGSMAW	VCYAVAQKLL	SAQFGPQIL	LLIYAASFAV
201	<u>FLIXAEPAHI</u>	<u>GSLDGTIAWV</u>	<u>CFVYCCNLTL</u>	IGYGSFGEAL	KHWEASKVSA
251	<u>VTTLLPVFTV</u>	<u>IFSLLGHYVM</u>	<u>PDTFAAPDMN</u>	<u>GLGYVGALVV</u>	<u>VGGAVTAAVG</u>
301	DRPFKRR*				

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

50	1	ATGGA AAAACC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTTGGCGGC
	51	GATGACGTGG	GGGACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGCC	GGCGCGCGTA
	151	TTGTTTGGTTT	TGCTGGCATT	GGGCGGGCGG	TCCCGCAAGC	GCGGGGATTT
	201	TTCTTGGCAT	TCATTACAGC	TGCTGCTGCT	CGGCGTGACG	GGCATTTCGG
55	251	CAAACCTTTG	GCTGATTGCC	CAAGGGCTGC	ATTATATTTT	GCCGACCCAG
	301	ACGCAGGTTT	TGTGGCAGAT	TTCGCGGTTT	ACGATGATTG	TGTGTCGGCT
	351	GTTGGTGTTT	AAAGACCGGA	tgaCTGCCGC	GCAGAAAATC	GGTTTGGTTT
	401	TGCTGCTtTG	CGGTTtTGCTT	ATGTTTTtta	ACGACAAATT	CGGCGAGTTG
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG	CGGCAGCGTA
60	501	TATGGCCTGG	TGTGTGTATG	CCGTGGCGCA	AAAGCTGCTG	TCCGCGCAAT
	551	TCGGGCCGCA	ACAGATFCTG	CTGTTGATTT	ATGCGGCaag	tgccgccGTG
	601	TTCCtgcggT	TTGccgaaCC	GGCACACATC	GGAAAGTTTg	acGGTACgTT
	651	GGCGTGGGTT	TGTTTTGTGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG

-253-

701 GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG  
 751 GTAACAACCT TGCTCCCGT GTTTACCGTA ATATTTCTT TGCTCGGGCA  
 801 TTATGTGATG CCTGATACTT TTGCCGCGCC GGATATGAAC GGTTCGGGT  
 851 ATGTCGGCGC ACTGGTCGTG GTCGGGGTG CCGTTACGGC GCGGTGGGG  
 901 GACAGGCCGT TCAAACGCC CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV  
 51 LFLVLLALGGR LPKRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT  
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLVGLL MFFNDKFGEL  
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV  
 201 FLPPFAEPAHI GSLDGT LAWV CFVYCCLNTL IGYGSFGEAL KHWEASKVSA  
 251 VTTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALV VGGAVTAAGV  
 301 DRPFKRR\*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15		10	20	30	40	50	60
	orf104-1.pep	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR					
	orf104ng-1	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR					
20		10	20	30	40	50	60
	orf104-1.pep	LPKRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPTTMIVVGVLVF					
	orf104ng-1	LPKRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPTTMIVVGVLVF					
25		70	80	90	100	110	120
	orf104-1.pep	LPKRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPTTMIVVGVLVF					
	orf104ng-1	LPKRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPTTMIVVGVLVF					
30		130	140	150	160	170	180
	orf104-1.pep	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL					
	orf104ng-1	KDRMTAAQKIGLVLLVGLLMMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL					
35		190	200	210	220	230	240
	orf104-1.pep	SAQFGPQQILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL					
	orf104ng-1	SAQFGPQQILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL					
40		250	260	270			
	orf104-1.pep	KHWEASKVSAVTTLLPVFTVIXXLLGHYVMPETFAAP					
	orf104ng-1	KHWEASKVSAVTTLLPVFTVIFSLGHYVMPDTFAAPDMNGLGYVGALVVGGAVTAAGV					
		250	260	270	280	290	300

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

45 gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306  
 Score = 237 bits (598), Expect = 8e-62  
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

50 Query: 30 QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFVDAPTLVXXXXXXXXXXXXXXXXXXXXP- 88  
 Q+P M WG+LPIA++QVL ++A T+VW P  
 Sbjct: 3 QQPLLGFTTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

55 Query: 89 --KRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPTTMIVVGVLVF 146  
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F  
 Sbjct: 63 LMKVRQYAW----IMLIGVIGLTSNLLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118

60 Query: 147 KDRMTAAQKIXXXXXXXXXXXMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206  
 K+++ QKI +FFND+F +GL Y+ GV+L G++ WV Y +AQKI+  
 Sbjct: 119 KEKLGILHQIGLGLLLIGLGLFFNDRFDAFAGLNQYSTGVILGVGALIVVAYGMAQKLM 178

65 Query: 207 SAQFGPQQILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL 266  
 +F QQILL++Y A F+P A+ + + L LA +CF+YCCLNTLIGYGS+ EAL  
 Sbjct: 179 LRKFNSQQILLMMYLGCAIAFMPMADFSQVQELT-PLALICFTYCCLNTLIGYGSYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVFTVIFSLGHYVMPDTFAAPDMN 306

W+ SKVS V TL+P+FT++FS + HY P FAAP++N  
 Sbjct: 238 NRWDVSKVSVVITLVPLFTILFISHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from  
 5 *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```

1  ATGGTAGCTC GTCGGGCTCA TAACCCGAAG GTCGTAGGTT CGAATCCTGT
10  51 .CCCGCAACC TAATTTCAAA CCCCTCGGTT CAATGCCGAG GG.GTTTTGT
    101 T.TTGCTGT TTCCTGTTTC CTGTTTCCTG CCGCTCCGT TTTTGGCCGG
    151 ATTTTCCTTC CGGCCGCAAT ATCGGAACGG CAGACCGCCG TCTGTTTGCG
    201 GTTGCAAATT CAGGCAGTTT GGCTACAATC TTCCGCATTG TCTTCAAGAA
    251 AGCCAACCAT GCCGACCGTC CGTTTACCG AATCCGTCAG CAAACAAGAC
15  301 CTTGATGCTC TGTTCGAGTG GGCAAAGCA AGTTACGGTG CAGAAAGTTG
    351 CTGGAACACG CTGTATCTGA ACGGTCysCC TTTGGGCAAC CTGTCGCCGG
    401 AATGGGTGGA ACGCGTsmmA AAAGACTGGG AGGCAGGCTG CyCGGAGTCT
    451 TCAGACGGCA TTTTCTGAA TgCGGACGGc TGgCctGATA TGGgCGGAcg
    501 cTTACAGCAC CTCGCCCTCG GTTGGCACTG TCGGGGGCTG TTGGACGgsT
20  551 GGCGCAACGA GTGTTTCGAC CTGACCGACG GCGGCGGCAA CCCCTTGTTC
    601 ACGCTCGaAc GCGCGyTTT mCGTCCTkTC GGA CTGCTCA GCCGCGCCGT
    651 CCATCTCAAC GGTCTGACCG AATCGGACGG CCGATGGCAT TTCTGGATAG
    701 GCAGGCGCAG TCCGCACAAA GCAGTCGATC CCAACAACT CGACAATACT
    751 rCCGCCGGCG GTGTTTCCGG CGGCGAAATG CCGTCTGAAG CCGTGTGTGG
25  801 CGAAAGCAGC GAAGAAGCCG GTTTGGATAA AACGTGcTT CCGCTCATCC
    851 CCCCAGGTATC GCAGCTGCAC AGCCTGCGCT CCGTCACCGG GGGTGTACAC
    901 AATGAAATCC TGTATGTATT CGATGCCGTC CTGCC...
  
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

1  MVARRAHNPK VVGSNFXPAT XFQTPRFNAE XVLXLPVSCF LFPAASVFGR
30  51 IFLPAAISER QTAVCLRLQI QAVWLQSSAL SSRKPTMPTV RFTESVSKQD
    101 LDALFEWAKA SYGAESCWKT LYLNXPGLN LSPWVERVX KDWEAGCXES
    151 SDGIFLNADG WPDMMGRLQH LALGWHCAGL LDGWRNECFD LTDGGGNPLF
    201 TLERAXRPX GLLSRAVHLN GLTESDGRWH FWIGRRSPHK AVDPNKLNT
    251 XAGGVSGGEM PSEAVCRESS EEAGLDKTL PLIRPVSQHL SLRSVSRGVH
35  301 NEILYVFDAV LP...
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

1  ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC
40  51 TCTGTTTCGAG TGGGCAAAAAG CAAGTTACGG TGCAGAAAGT TGCTGGAAAA
    101 CGCTGTATCT GAACGGTCTG CCTTTGGGCA ACCTGTCGCC GGAATGGGTG
    151 GAACGCGTCA AAAAAGACTG GGAGGCAGGC TGCTCGGAGT CTTACAGACGG
    201 CATTTTTCTG AATGCGGACG GCTGGCCTGA TATGGGCGGA CGCTTACAGC
    251 ACCTCGCCCT CGGTTGGCAC TGTGCGGGGC TGTGGACGCG CTGGCGCAAC
    301 GAGTGTTCG ACCTGACCGA CGGCGGCGGC AACCCTTGT TCACGCTCGA
    351 ACGCGCGGCT TTCCGTCCTT TCGGACTGCT CAGCCGCGCC GTCCATCTCA
45  401 ACGGTCTGAC CGAATCGGAC GGCCGATGGC ATTTCTGGAT AGGCAGGCGC
    451 AGTCCGCACA AAGCAGTCGA TCCCAACAAA CTCGACAATA CTGCCCGCGG
    501 CCGTGTTCG GCGGCGGAAA TGCCGTCTGA AGCCGTGTGT CGCGAAAGCA
    551 GCGAAGAAGC CGGTTGGGAT AAAACGCTGC TTCCGCTCAT CCGCCCGGTA
    601 TCGCAGCTGC ACAGCCTGCG CTCCGTCAGC CGGGTGTGAC ACAATGAAAT
50  651 CCTGTATGTA TTCGATGCCG TCCTGCCCGA AACCTTCCTG CCTGAAAATC
    701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CCGTCTGTTG
    751 GATGCCATGT TGTGCGGAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
    801 GGACGCGTTT TGCCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
    851 AGTGGCTGGA CGGCATACGT TTATAG
  
```

55 This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

-255-

1 MPTVRFTESV SKQDLDALE WAKASYGAES CWKTLYLNL PLGNLSPEWV  
 51 ERVKKDWEAG CSESSDGIFL NADGWPMGG RLQHLALGWH CAGLLDGWRN  
 101 ECFDLTDGGG NPLFTLRAA FRPGLLSRA VHLNGLTESD GRWHEWIGRR  
 151 SPHKAVDPNK LDNTAAGVSV GGEMPSEAVC RESSEEAGLD KTLPLIRPV  
 201 SQLHSLRSVS RGVHNEILYV FDAVLPEFEL PENQDGEVAG FEKMDIGLL  
 251 DAMLSGNMMH DAQLVTLDFA CRYGLIDAAH PLSEWLDGIR L\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N.*

10 *meningitidis*:

	60	70	80	90	100	110
orf105.pep	ISERQTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAES					
orf105a				MPTVRFTESVSKHDLDALEWAKASYGAES		
				10	20	30
	120	130	140	150	160	170
orf105.pep	CWKTLYLNGXPLGNLSPEWVERVVKDWEAGCXESSDGIFLNADGWPMGGRLQHLALGWH					
orf105a	CWKTLYLNLPLGNLSPEWAERVVKDWEAGCXESSDGIFLNADGWPMGGRLQHLARIWK					
	40	50	60	70	80	90
	180	190	200	210	220	230
orf105.pep	CAGLLDGWRNECFDLTDGGGNPLFTLERAXXRPXGLLSRAVHLNGLTESDGRWHEWIGRR					
orf105a	EAGLLHGWRDECFDLTDGGSNPLFALERAAFPPGLLSRAVHLNGLVESDGRWHEWIGRR					
	100	110	120	130	140	150
	240	250	260	270	280	290
orf105.pep	SPHKAVDPNKLDNTXAGVSGGEMPSEAVCRESSEEAGLDKTLPLIRPVSQHLHSLRSVS					
orf105a	SPHKAVDPDKLDNTAAGVSSGELPSETVCRESSEEAGLDKTLPLIRPVSQHLHSLRPVS					
	160	170	180	190	200	210
	300	310				
orf105.pep	RGVHNEILYVFDVLP					
orf105a	RGVHNEILYVFDVLPETFLPENQDGEVAGFEKMDIGLLAAMLSGNMMHDAQLVTLDFA					
	220	230	240	250	260	270

40 The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

1 ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACACG ACCTTGATGC  
 51 CCTATTCGAG TGGGCAAAGG CAAGTTACGG TCGGGAAGT TGCTGGAAAA  
 101 CGCTGTATCT GAACGGTCTG CTTTGGGCA ATCTGTCGCC GGAATGGGCG  
 151 GAGCGCGTCA AAAAAGACTG GGAGGCAGGC TGCTCGGAGT CTTGAGACGG  
 201 CATTTTCCTG AATGCGGACG GCTGGCCAGA TATGGGCAGA CGCTTGCAGC  
 251 ACCTCGCCCG AATATGGAAA GAAGCGGGAC TGCTTCACCG CTGGCGCGAC  
 301 GAGTGTTCG ACCTGACCGA CGGCGGCAGC AATCCCTTGT TCGCGCTCGA  
 351 ACGCGCCGCT TTCCGTCCGT TCGGACTGCT CAGCGCGGCC GTCCATCTCA  
 401 ACGGTTTGGT CGAATCGGAC GGCGGATGGC ATTTCTGGAT AGGCAGCGCG  
 451 AGTCCGCACA AAGCAGTCGA TCCGCACAAA CTCGACAATA CTGCCGCCCG  
 501 CGGTGTTTCC AGCGGTGAAT TGCCGTCTGA AACCGTGTGT CGCGAAAGCA  
 551 CGGAAGAAGC CGGTTGGAT AAAACGCTGC TTCCGCTCAT CCGCCCGTA  
 601 TCGCAGCTGC ACAGCCTGCG CCCCCTCAGC CGGGGTGTGC ACAATGAAAT  
 651 CCTGTATGTA TTCGATGCCG TCCTGCCCCA AACCTTCCTG CCTGAAATC  
 701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CGGTCTGTGT  
 751 GCTGCCATGT TGTGGGAAA CATGATGCAC GACGCGCAAC TGGTACGCT  
 801 GGACGCGTTT TGCCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG  
 851 AGTGGCTGGA CGGCATACGT TTATAG

This encodes a protein having amino acid sequence <SEQ ID 414>:

60 1 MPTVRFTESV SKHDLDALE WAKASYGAES CWKTLYLNL PLGNLSPEWA  
 51 ERVKKDWEAG CSESSDGIFL NADGWPMGR RLQHLARIWK EAGLLHGWRD

-256-

```

101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR
151 SPHKAVDPDK LDNTAAGGVS SGELPSETVC RESSEEAGLD KTLPLIRPV
201 SQLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
251 AAMLSGNMMH DAQLVTLDFA CRYGLIDAAH PLSEWLDGIR L*

```

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

```

10 orf105a.pep      10      20      30      40      50      60
    MPTVRFTESVSKHDLDALEWAKASYGAESCWKTLYLNLPLGNLSPEWAEVVKKDWEAG
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105-1      10      20      30      40      50      60
    MPTVRFTESVSKQDLDALEWAKASYGAESCWKTLYLNLPLGNLSPEWAEVVKKDWEAG

15 orf105a.pep      70      80      90      100     110     120
    CSESSDGIFLNADGWPDMMGRRLQHLARIWKEAGLLHGWRDECFLTDGGSNPLFALERAA
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105-1      70      80      90      100     110     120
    CSESSDGIFLNADGWPDMMGRRLQHLALGWHCAGLLDGWRNECFDLTDGGSNPLFTLERAA

20 orf105a.pep      130     140     150     160     170     180
    FRPFGLLSRAVHLNGLVESDGRWHFWIGRRSPHKAVDPDKLDNTAAGGVSSGELPSETVC
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105-1      130     140     150     160     170     180
    FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPNKLNTAAGGVSSGEMPSEAVC

25 orf105a.pep      190     200     210     220     230     240
    RESSEEAGLDKTLPLIRPVSQHLSLRPVSRGVHNEILYVFDVLPETFLPENQDGEVAG
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105-1      190     200     210     220     230     240
    RESSEEAGLDKTLPLIRPVSQHLSLRSVSRGVHNEILYVFDVLPETFLPENQDGEVAG

30 orf105a.pep      250     260     270     280     290
    FEKMDIGGLLAAMLSGNMMHDAQLVTLDFAFCRYGLIDAAHPLSEWLDGIRLX
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105-1      250     260     270     280     290
    FEKMDIGGLLDAMLSGNMMHDAQLVTLDFAFCRYGLIDAAHPLSEWLDGIRLX

35

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N. gonorrhoeae*:

```

40 orf105.pep      MVARRAHNPKVVGSNPXPATXFQTPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAAISER      60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105ng      MVARRAHNPKVVGSNPAPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAAISER      55

45 orf105.pep      QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAESCWK      120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105ng      QAAVCLRLQIQAVWLQSSALCSRKPAMPPTVRFTESVSKQDLDALEWAKASYGAESCWK      115

50 orf105.pep      LYLNGXPLGNLSPEWVERVXKDWEAGCXESSDGIFLNADGWPDMMGRRLQHLALGWHCAGL      180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105ng      LYLNRLPLGNLSPEWAERIKKDWEAGCSESSNGIFLNADGWPDMMGRRLQHLARTWNKAGL      175

55 orf105.pep      LDGWRNECFDLTDGGGNPLFTLERAXRXPGLLSRAVHLNGLTESDGRWHFWIGRRSPHK      240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105ng      LHGWRNECFDLTDGGGNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK      235

60 orf105.pep      AVDPNKLNTXAGGVSGGEMPSEAVCRSSEEAGLDKTLPLIRPVSQHLSLRSVSRGVH      300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105ng      AVDPGKLDNIAGGVSGGEMPSEAVCRSSEEAGLDKTLFPLIRPVSRLHSLRPVSRGVH      295

orf105.pep      NEILYVFDVLP      312
    |||||:|||||
orf105ng      NEILYVFDVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVTLDFAFYRG      355

```

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

-257-

1 MVARRAHNPK VVGSNPAPAT KYQTERFNAE GVLFFLFPA SVFCRIFLPA  
 51 AISERQAAVC LRLQIQAVWL QSSALCSRKP AMPTVRFTES VSKQDLDALE  
 101 ERAKASYGAE SCWKTLYLNR LPLGNLSPEW AERIKKDWEA GCSESSNGIF  
 151 LNADGWPDGM GRLQHLARTW NKAGLLHGWR NECFDLTDGG GNPLFTLERA  
 201 AFRPFGLLIR AVHLNGLVES NGRWHFWIGR RSPHKAVDPG KLDNIAGGGV  
 251 SGGEMPSEAV CRESSEEAGL DKTFLPLIRP VSRHLSLRPV SRGVHNEILY  
 301 VFDAVLPETF LPENQDGEVA GFEMDIGGL LDAMLSKNMM HDAQLVTLDA  
 351 FYRYGLIDAA HPLSEWLDGI RL\*

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

1 ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC  
 51 CCTGTTTCGAG CGGGCAAAAAG CAAGTTACGG TGCCGAAAGT TGCTGAAAAA  
 101 CGCTGTATCT GAACCGTCTT CCTTTGGGCA ATCTGTGCGC GGAATGGGCT  
 151 GAGCGCATCA AAAAAGACTG GGAGGCAGGC TGCTCCGAGT CTTCAGACGG  
 201 CATTTTTCTG AATGCGGACG GCTGGCCGGA TATGGGCGGA CGCTTGACAG  
 15 ACCTCGCCCG CACATGGAAC AAGGCGGGGC TGCTTCACGG ATGGCGCAAC  
 301 GAGTGTTCG ACCTGACCGA CGGCGGCGGC AACCCTTGT TCACGCTCGA  
 351 ACGCGCCGCT TTCCGTCCGT TCGGACTACT CAGCCGCGCC GTCCATCTCA  
 401 ACGGTTTGGT CGAATCGAAC GGCAGATGGC ATTTTGGAT AGGCAGGCGC  
 451 AGTCCGCACA AAGCAGTCGA TCCCGCAAG CTCGACAATA TTGCCGCGCG  
 20 501 CGGTGTTTCC GCGGCGGAAA TGCCGTCTGA AGCCGTGTGC CGCGAAAGCA  
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCAGTA  
 601 TCGCGGCTGC ACAGCCTTCG CCCCCTCAGC CGAGGTGTGC ACAATGAAAT  
 651 CCTGTATGTG TTCGATGCCG TCCTGCCCGA AACCTTCCTG CCTGAAAAATC  
 701 AGGATGGCGA GGTAGCGGGT TTTGAAAAGA TGGACATTGG CGGCCTATTG  
 25 751 GATGCCATGT TGTGAAAAA CATGATGCAC GACGCGCAAC TGGTTACGCT  
 801 GGACGCGTTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG  
 851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

1 MPTVRFTESV SKQDLDALE RAKASYGAES CWKTLYLNR LPLGNLSPEWA  
 30 51 ERIKKDWEAG CSESSDGIFL NADGWPDMMG RLQHLARTWN KAGLLHGWRN  
 101 ECFDLTDGGG NPLFTLERA FRPFGLLSRA VHLNGLVESN GRWHFWIGRR  
 151 SPHKAVDPGK LDNIAGGGVS GGEMPSEAVC RESSEEAGLD KTLFPLIRPV  
 201 SRLHSLRPVS RGVHNEILYV FDAVLPETF PENQDGEVAG FEKMDIGGLL  
 251 DAMLSKNMMH DAQLVTLDAF YRYGLIDAAH PLSEWLDGIR L\*

35 ORF105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
orf105-1.pep		MPTVRFTESVSKQDLDALE	FEWAKASYGAES	SCWKTLYLNL	LPLGNLSPEW	VERVKKDWEAG	
40 orf105ng-1		MPTVRFTESVSKQDLDALE	FEWAKASYGAES	SCWKTLYLNR	LPLGNLSPEW	AERIKKDWEAG	
		10	20	30	40	50	60
		70	80	90	100	110	120
orf105-1.pep		CSESSDGIFLNADGWPDMMG	RLQHLALGWH	CAGLLDGRNE	CFDLTDGG	GNPLFTLERA	
45 orf105ng-1		CSESSDGIFLNADGWPDMMG	RLQHLARTWN	KAGLLHGWRNE	CFDLTDGG	GNPLFTLERA	
		70	80	90	100	110	120
		130	140	150	160	170	180
orf105-1.pep		FRPFGLLSRAVHLNGLTES	DGRWHFWIGRR	SPHKAVDPN	KLDNTAAG	GVSGGEMPSEAVC	
50 orf105ng-1		FRPFGLLSRAVHLNGLVES	NGRWHFWIGRR	SPHKAVDPG	KLDNIAGG	GVSGGEMPSEAVC	
		130	140	150	160	170	180
		190	200	210	220	230	240
orf105-1.pep		RESSEEAGLDKTLPLIRPV	SQLHSLRSV	SRGVHNEILY	VFDAVLPET	FLPENQDGEVAG	
55 orf105ng-1		RESSEEAGLDKTLFPLIRPV	SRLHSLRPV	SRGVHNEILY	VFDAVLPET	FLPENQDGEVAG	
		190	200	210	220	230	240
		250	260	270	280	290	
orf105-1.pep		FEKMDIGGLLDAMLSKNMMH	DAQLVTLDA	FCRYGLIDAA	HPLSEWLDGIR	LX	
60 orf105ng-1		FEKMDIGGLLDAMLSKNMMH	DAQLVTLDA	FYRYGLIDAA	HPLSEWLDGIR	LX	
		250	260	270	280	290	

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

5  sp|P41888|TNR3_SCHPO_THIAMIN_PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
   >gi|1076928|pir||S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
   (Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
   [Schizosaccharomyces pombe] >gi|2330852|gnl|PID|e334056 (Z98533) thiamin
   pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
   Score = 105 bits (259), Expect = 4e-22
   Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

10  Query: 268 NKAGLLHGWNECFDLTDGGGNPLFTLERAARFPFGLLSRAVHLNGLVESNGRW--HFWI 441
      N G+ WRNE + + P+ +ER F FG LS VH + + W+
   Sbjct: 96 NTFGIADQWRNELYTVYGSKKPVLAVERGGFWLFGFLSTGVHCTMYIPATKEHPLRIWV 155

15  Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESSEEAGLDKTLFPLIRPVSRLHSLR 621
      RRSP K P LDN GG++ G+ + +E SEEA LD + LI P + ++
   Sbjct: 156 PRRSPTKQTPWNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNLICPGTVSYIK 214

   Query: 622 PVSRG-VHNEILYVFDVAVLPETFLPENQDGEVAGFEKMDIGLLDAMLSKNMMHDAQLVT 798
      R + E+ YVED + + +P DGEVAGF + + +L + K+ + LV
20  Sbjct: 215 MEKRHWIQPELQYVFDLPVDDLVIPIRINDGEVAGFSLPLNQVLHELELKSFKPNCALVL 274

   Query: 799 LDAFYRYGLIDAAHP 843
      LD R+G+I HP
   Sbjct: 275 LDFLIRHGIITPQHP 289

```

- 25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 30 419>:

```

1  ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
51  CCAAACAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
101 CCCTATGGAC GACATTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
151 TTGATATTTG GTAACATATC GCGAAAGACA ACAGTGGAGG GACAAATTTT
35 201 ACCTGCATCG GCGGTAATCA GGGTGTATGC ACCGgATACG rGkACAATTA
251 CAGCGAAATT CGTGGAAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
301 TTTGCGCTTT CGACCTCACG TTTGCGCGCA GGAGGTAGCG TGCAGCAGCA
351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAAGTGG
40 401 GTCGTCTGAA GCTGATACAC GGAATGAAA CGCGCAGCcT TAAAGCAACT
451 GTCGAACGTT TGGAAGAACCA GGAAGTCCAT ATTTGCGCAAC AGATAGACGG
501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

45 1  MNRPKQPFER PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51 51  LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
101 101 FALSTSFRFGA GGSVQQQLKT EAVLKKTLE QELGRLKLIH GNETRSLKAT
151 151 VERLENQELH ISQQIDGQKR RIRLAEEMLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

- 50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

10 20 30 40 50 60



-259-

5	orf107.pep	MNRPKQPFPRPEVAVARQTS	LTGKVLTRPLSFSLWTT	FASISALLIILFLIFGNYTRKT	
	orf107a	MNRPKQPFPRPEVAVARQTS	LTGKVLTRPLSFSLWTT	FASISALLIILFLIFGNYTRKT	
10		10 20 30 40 50 60			
		70 80 90 100 110 120			
15	orf107.pep	TVEGQILPASGVIRVYAPDTXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT			
	orf107a	TVEGQILPASGVIRVYAPDTGTITAKFXEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT			
20		70 80 90 100 110 120			
		130 140 150 160 170 180			
25	orf107.pep	EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLQ			
	orf107a	EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLQ			
30		130 140 150 160 170 180			
		189			
35	orf107.pep	KYRFLSXQX			
	orf107a	KYRFLSANDAVPKQEMMNVAELLEQKAKLDAYRREEVGLLQEIRTQNLTLXSLPQAAX			
40		190 200 210 220 230			

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

25	1	ATGAATAGAC	CCAAGCAACC	NTTCTTCCGT	CCCGAAGTCG	CCGTTGCCCCG
	51	CCAAACCAGC	CTGACGGGTA	AAGTGATTCT	GACACGACCG	TTGTCATTTT
30	101	CCCTATGGAC	GACATTTGCA	TCGATATCTG	CGTTATTGAT	TATCCTGTTT
	151	TTGATATTTG	GTAACATATAC	GCGAAAGACA	ACAGTGGAGG	GACAAATTTT
35	201	ACCTGCATCG	GGCGTAATCA	GGGTGTATGC	ACCGGATACG	GGGACAATTA
	251	CNGCGAAATT	CNTGGAAGAT	GGAGAAAAGG	TTAAGGCTGG	CGACAAGCTA
40	301	TTTGCCTTTT	CGACCTCACG	TTTCGGCGCA	GGAGATAGCG	TGCAGCAGCA
	351	GTTGAAAACG	GAGGCAGTTT	TGAAGAAAAC	GTTGGCAGAA	CAGGAAGTGG
45	401	GTCGTCTGAA	GCTGATACAC	GGGAATGAAA	CGCGCAGCCT	TAAAGCAACT
	451	GTCGAACGTT	TGGAAAACCA	GGAATCCAT	ATTTCGCAAC	AGATAGACGG
50	501	TCAGAAAAGG	CGCATTAGAC	TTGCGGAAGA	AATGTTGCAG	AAATATCGTT
	551	TCCTATCCGC	CAATGATGCA	GTGCCAAAAC	AAGAAATGAT	GAATGTCAAG
55	601	GCAGAGCTTT	TAGAGCAGAA	AGCCAACTT	GATGCCTACC	GCCGAGAAGA
	651	AGTCGGGCTG	CTTCAGGAAA	TCCGCACGCA	GAATCTGACA	TTGGNNAGCC
60	701	TCCCCAAGC	GGCATGA			

This encodes a protein having amino acid sequence <SEQ ID 422>:

40	1	MNRPKQPFPR	PEVAVARQTS	LTGKVLTRP	LSFSLWTTFA	SISALLIILF
	51	LIFGNYTRKT	TVEGQILPAS	GVIRVYAPDT	GTITAKFXED	GEKVKAGDKL
45	101	FALSTSRFGA	GDSVQQQLKT	EAVLKKTAE	QELGRLKLIH	GNETRSLKAT
	151	VERLENQELH	ISQQIDGQKR	RIRLAEEMLQ	KYRFLSANDA	VPKQEMMNVK
50	201	AELLEQKAKL	DAYRREEVGL	LQEIRTQNL	LXSLPQAA*	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N. gonorrhoeae*:

50	orf107.pep	MNRPKQPFPRPEVAVARQTS	LTGKVLTRPLSFSLWTT	FASISALLIILFLIFGNYTRKT	60
	orf107ng	MNRPKQPFPRPEVAIARQTS	LTGKVLTRPLSFSLWTT	FASISALLIILFLIFGNYTRKT	60
55	orf107.pep	TVEGQILPASGVIRVYAPDTXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT			120
	orf107ng	TMEGQILPASGVIRVYAPDTGTITAKFVEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT			120
60	orf107.pep	EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLQ			180
	orf107ng	EAVLKKTAEQELGRLKLIHENETRSLKATVERLENQKLHISQQIDGQKRRIRLAEEMLR			180
65	orf107.pep	KYRFLSXQ	188		
	orf107ng	KYRFLSAQ	188		

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

1  MNRPKQPFER PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
5  51  LIFGNYTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKKTAE QELGRLKLIH ENETRSLKAT
151 VERLENQKLH ISQQIDGQKR RIRLAEEMLR KYRFLSAQ*

```

Based on the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC
15 51  GTGCGGCAAA TCCGTAAATA CGGCGGTACA GCCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAAGTCATAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GCGCACTTGG AAGCCGTCAG CGGCAAATGT
20 301 ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

25 1  MLNTFFAVLG GCLLXLPCGK SVNTAVQPQN AVQSAPKPVF KVIYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGGKLT DYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

30 1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCCGC
51  CTGCGGCAAA TCCGAAAATA CGGCGGAACA GCCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAAGTCAAAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
35 201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GCGCACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
40 451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

45 1  MLKTSFAVLG GCLLLAACGK SENTAEQPQN AVQSAPKPVF KVKYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGGKLT DYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181 aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

5	orf108.pep	MLNTFFAVLGGCLLXPCGKSVNTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE	60
	orf108ng	MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVIYIDNTAIAGLALGQSSE	60
10	orf108.pep	GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT	120
	orf108ng	GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120
15	orf108.pep	LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH	181
	orf108ng	LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH	181

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

20	orf108-1.pep	MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE	60
	orf108ng-1	MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVIYIDNTAIAGLALGQSSE	60
25	orf108-1.pep	GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT	120
	orf108ng-1	GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120
30	orf108-1.pep	LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH	181
	orf108ng-1	LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH	181

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

30	1	ATGCTGAAa	tacctTTTGC	CGTGtgggc	ggCtgcctGC	TGCTTGCCGC
	51	CTGCGGCAAA	TCCGAAAATa	cggcggaACA	GCCGCAAAAT	gcggCACAA
	101	GCGCGCCGAA	ACCGGTTTTc	AAAGTCAAAT	ACATCGACAA	TACGGCGATT
	151	GCCGGTTTGG	CTTTGGGACA	AAGTAGCGAA	GGCAAACCA	acgacgGCAA
35	201	AAAACAAATC	AGTTATccgA	TTAAAGGCTT	GCCGGAACAA	Aacgcgctcc
	251	gGCTGACCGG	AAAGCATCCC	AACGACTTGG	AagccgtcgT	CGGCAAATGT
	301	ATGGAAACCG	ACGGAAAGGA	CGCGCCTTCG	GGCTGGGCGG	AAAACGGCGT
	351	GTGCCATACC	TTGTTTGCCA	AACTGGTGCG	CAATATCGCC	GAAGACGGCG
40	401	GCAAACTGAC	TGATTACCTG	ATTTCGCATT	CCGCCCTGCA	ACCCTATCAG
	451	GCAGGCAAAA	GCGGCTATGC	CGCCGTGCAG	AACGGACGCT	ATGTGCTGGA
	501	AATCGACAGC	GagggGGCGT	TTTATttccg	cgcgcgcat	tattgA

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

1	MLKIPFAVLG	GCLLLAACGK	SENTAEQPQ	NAASAPKPV	FKVIYIDNTAI
51	AGLALGQSSE	GKTNDGKKQI	SYPIKGLPEQ	NAVRLTGKHP	NDLEAVVGK
101	METDGKDAPS	GWAENGVCHT	LFAKLVGNI	AEDGGKLT	DYLVSHAALQPYQ
151	AGKSGYAAVQ	NGRYVLEIDS	EGAFYFRRRH	Y*	

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**50 Example 51**

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

-262-

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTGTTGGTTG CGATGATTGC  
 51 CGGATTTATC GATgcatTg cGggCGGGGG TGGTTTGATT ACGCTGCCCCG  
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG  
 151 CTGCAAgCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA  
 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG  
 251 TAGGCGGCGT GGCCTGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT  
 301 CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT  
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT  
 401 TTTTCTGTT cGGGCTGACG GTCGC.ACCG CTTTGGGTT TTTACGACGG  
 10 451 TGTGTTCCGA CCGGGTGTGCG GCTCGTTTTT TCTGATTGCC TTTATTGTTT  
 501 TGCTCGGCTG CAAGCTGTTG AACCGGATGT CTACACCAA ATTGGCGAAC  
 551 GTTGCCCTGCA ATCTTGGTTC GCTATCGGTA TTCCTGCTGC ACGGTTTCGAT  
 601 TATTTTCCCG ATTGCGGCAA CGaTGCGGT CGGTGCGTTC GTCGGtGCCA  
 651 ATTTAgGTGC GAGATTTGCC GTaCgctTCG GTTCGAAGCT GATTAA

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK  
 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI  
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VXTAFGFLRR  
 151 CVRTGCRLVF SDCLYCFARL QAVRDVLHQ IGERCLQSWF AIGIPAARFD  
 20 201 YFPDCGNDGG RCVCRCEFRG EICRTLRFEA D\*

Further work revealed the following DNA sequence <SEQ ID 433>:

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTGTTGGTTG CGATGATTGC  
 51 CGGATTTATC GATGCGATTG CCGGCGGGGG TGGTTTGATT ACGCTGCCCCG  
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG  
 151 CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA  
 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG  
 251 TAGGCGGCGT GGCCTGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT  
 301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT  
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT  
 401 TTTTCTGTT CCGGCTGACG GTCGCACCGC TTTTGGGTTT TTACGACGGT  
 10 451 GTGTTCCGAC CCGGTGTGCG CTCGTTTTTT CTGATTGCCT TTATTGTTTT  
 501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG  
 551 TTGCCTGCAA TCTTGGTTTC CTATCGGTAT TCCTGCTGCA CGGTTCGATT  
 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA  
 15 651 TTTAGGTGCG AGATTTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC  
 701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAATTGTT GATAGACGAG  
 751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK  
 40 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI  
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG  
 151 VFGPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLHGS  
 201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE  
 251 RNPLYQMIVS MF\*

45 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
50	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNK	LQAAAATFSA				
	orf109a	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNK	LQAAAATFSA				
		10	20	30	40	50	60
55	orf109.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
	orf109a	TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120

-263-

```

              130      140      150      160      170      180
orf109.pep    KLDGSKEGKARMSFFLFGLTVXTAFGFLRRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ
5             |||||
orf109a       KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
              130      140      150      160      170      180

```

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

```

1  ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTTGGTTG CGATGATTGC
51 CGGATTATATC GATGCGATTG CGGGTGGGGG TGGTTTGATT ACGCTGCCTG
10 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TGCACGCAA
201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCGGCA GCATCGTTTG
251 CAGGCGGCGT GGTCCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
15 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401 TTTTCTGTT CGGTCTGACG GTTGCAACAC TTTTGGGTTT TTACGACGGT
451 GTGTTCCGAC CGGGTGTGCG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG
551 TTGCCGTCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
20 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
651 TTTAGGTGCG AGATTGCGCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAATTGTT GATAGACGAG
751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 436>:

```

25 1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151 VFGPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLLHGS I
201 IFPIAATMAV GAFVGANLGA REAVRFGSKL IKPLLIVISI SMAVKLLIDE
30 251 RNPLYQMIVS MF*

```

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

```

              10      20      30      40      50      60
orf109a.pep    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
35 orf109-1     MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
              10      20      30      40      50      60

              70      80      90      100     110     120
orf109a.pep    TVSFARKGLIDWKKGLPIAAASFAGGVGALSLSVSLVSKDILLAVVPVLLIFVALYFVFSP
40 orf109-1     TVSFARKGLIDWKKGLPIAAASFVGGVAGALSLSVSLVSKDILLAVVPVLLIFVALYFVFSP
              70      80      90      100     110     120

              130     140     150     160     170     180
45 orf109a.pep    KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
orf109-1        KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
              130     140     150     160     170     180

              190     200     210     220     230     240
50 orf109a.pep    LANVACNLGSLSVFLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
orf109-1        LANVACNLGSLSVFLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
              190     200     210     220     230     240

55              250     260
orf109a.pep    SMAVKLLIDERNPLYQMIVSMFX
orf109-1        SMAVKLLIDERNPLYQMIVSMFX
60              250     260

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N.gonorrhoeae*:

```

5      orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60
      orf109ng   MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60

      orf109.pep  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120
10     orf109ng   TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120

      orf109.pep  KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVRDVLHQ  180
      orf109ng   KLDGSKEGKARMSFFLFGLTVATAFGFLRRCVRTGCRLVFSDCLYCFARLQAVRDVLHQ  180

15     orf109.pep  IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVRCCEFRCEICRPLRFEAD  231
      orf109ng   IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVRCCEFRCEICRPLRFEAD  231

```

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino acid sequence <SEQ ID 438>:

```

1      MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51     LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101    LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VATAFGFLRR
151    CVRTGCRLVF SDCLYCFARL QAVRDVLHQ IGERCLQSWF AIGIPAARFD
25     201     YFPDCGNDGG RCVRCCEFR C EICRPLRFEA D*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

```

1      ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTCGGTTC CGATGATCGC
51     CGGATTTATC GATCGGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCTG
101    CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
30     151    CTGCAAGCAG CCGCTGCTAC GTTTCGGCT ACGGTTTCTT TTGCACGCAA
201    AGGTTTGATT GATTGGAAGA AAGTCTCCC GATTGCCGCA GCATCGTTTG
251    CAGGCGGCGT GGTGCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301    TTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGCG TGTATTTTGT
351    GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401    TTTTCTTATT CGGGCTGACG GTTGCAACGC TTTTGGGTTT TTACGACGGT
35     451    GTGTCGGAC CGGTGTGCG CTCGTTTTT CTGATTGCCT TTATGTTTTT
501    GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG
551    TTGCTTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTCGATT
601    ATTTTCCCGA TTGTGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
40     651    TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701    TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
751    AGAAATCCGC TGTATCAGAT GATTGTTTTG ATGTTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

```

1      MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
45     51     LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101    LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151    VFGPGVGSFF LIAFIVLLGC KLLNAMS YK LANVACNLGS LSVLLHGS I
201    IFPIVATMAV GAFVGANLGA RFAVRFGSKL IKPLLVISI SMAVKLLIDE
251    RNPLYQMIVS MF*

```

ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

```

10      20      30      40      50      60
orf109ng-1.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
55     orf109-1    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
10      20      30      40      50      60

70      80      90      100     110     120
orf109ng-1.pep TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP

```

-265-

```

      |||
orf109-1 TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70      80      90      100     110     120

5      130      140      150      160      170      180
orf109ng-1.pep KLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      |||
orf109-1 KLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      130      140      150      160      170      180

10     190      200      210      220      230      240
orf109ng-1.pep LANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      |||
orf109-1 LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      190      200      210      220      230      240

15     250      260
orf109ng-1.pep SMAVKLLIDERNPLYQMIVSMFX
      |||
20     orf109-1 SMAVKLLIDERNPLYQMIVSMFX
      250      260

```

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```

25     sp|P29942|YCB9_PSEDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
>gi|94984|pir|I38164 hypothetical protein 9 - Pseudomonas sp >gi|551929
(M62866) ORF9 [Pseudomonas denitrificans] Length = 261
Score = 175 bits (439), Expect = 3e-43
Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

30     Query: 41 PPVSAIATNKLQXXXXXXXXXXXXXRRKGLIDWKKGLPIXXXXXXXXXXXXXXXXXXXXKDI 100
      PP+ + TNKLQ R+G ++ K+ LP+ D+
      Sbjct: 43 PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMALMSAAGAVLGALLATIVPGDV 102

      Query: 101 LLAVVPVLLIFVALYFVFSPKLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFF 160
      L A++P LLI +ALYF P + G + +R++ F+F LT+ PL+GFYDGVFGPG GSFF
35     Sbjct: 103 LKAILPFLIIAIALYFGLKPNM-GDVDQHSRVTPFVFTLTPLVPLIGFYDGVFGPGTGSFF 161

      Query: 161 LIAFIVLLGCKLLNAMS YTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGA 220
      ++ F+ L G +L A ++TK N N+G+ VFL G++++ + M +G F+GA +G+
40     Sbjct: 162 MLGFVTLAGFVLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLLMGLGQFLGAQVGS 221

      Query: 221 RFAVRFGSKLIKPLLIVISISMAVKLLIDERNPL 254
      R+A+ G+K+IKPLL+++SI++A++LL D +PL
      Sbjct: 222 RYAMAKGAKIIPLLVIVSIALAIRLLADPTHPL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

```

50      1 ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACCAG
      51 CCGCAGACGG ATTATTTGGT CAAATTCGGA TCGTTTGGG CGAG.ATTTT
      101 TGGTTTCTG GGA CTGTATG ACGTCTATG TTCGGCATGG TTTGTCGTTA
      151 TCATGATGTT TTTGGTGGTT TCTACCAGT TGTGCCTGAT TCGCAATGTG
      201 CCGCCGTTCT GGC GCGAAAT GAAGTCTTTT CGGGAAAAGG TTAAAGAAAA
55     251 ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGGATGTA AAAATTGCGC
      301 CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTTC AAGGAAAACC
      351 ATTAACCGTG AAGACGGGTC GGTCTGATT GCCGCCAAAA AAGGCACAAT
      401 GAACAAATGG GGCTATATCT TTGCCCATGT TGCTTTGATT GTCATTTGCC
      451 TGGGCGGGTT GATAGACAGT AACCTGCTGT TGAAACTGGG TATGCTGACC
60     501 GGTCCGATTG TTCCGACAA TCAGGCGGTT TATGCCAAGG ATTTTC.AAGC

```

-266-

551 CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTAGGG GCAACGTCAA  
 601 TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTTCCTGA

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

5           1   ..LLGIASVIGT LLQONQPQTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI  
           51   MMFLVVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP  
          101   EVAKRYLEVQ GFQGTINRE DGSVLIAAKK GTMNKGYIF AHVALIVICL  
          151   GGLIDSNLLL KLGMLTGRIF RTIRRFMPRI XKPESXFGCV QSLI\*GQRQY  
          201   FXRGRVRMWF S\*

Computer analysis of this amino acid sequence gave the following results:

#### 10   Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf88a.pep	MSKSRRSP	LLSRPWF	FAFFSSMR	FAVALLS	LLGIASVIGT	VLQONQPQTDYLVKFGSFWA
15	orf110					LLGIASVIGT	LLQONQPQTDYLVKFGSFWA
					10	20	30
		70	80	90	100	110	120
20	orf88a.pep	QIFGFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNVPPFW	REMKSFREKVKEKSLAAMRH			
	orf110	XIFGFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNVPPFW	REMKSFREKVKEKSLAAMRH			
		40	50	60	70	80	90
25	orf88a.pep	SSLLDVKIAPEVAKRYLEVQ	GFQGTINREDGSVLIAAKK	GTMNKGYIFAHVALIVICL			
	orf110	SSLLDVKIAPEVAKRYLEVQ	GFQGTINREDGSVLIAAKK	GTMNKGYIFAHVALIVICL			
		100	110	120	130	140	150
30	orf88a.pep	GGLIDSNLLLKL	GMLTGRIVPDNQAVYAKDF	KPESILGASNLSFRGNVNISEG	QSADVVVF		
	orf110	GGLIDSNLLLKL	GMLTGRIFRTIRRFMPRI	XKPESXFGCVQSLIXGQRQY	FXRGRVRMWF		
35		160	170	180	190	200	210
	orf88a.pep	LNADNGILVQDLP	FEVKLKKFHIDFYNTGMPR	DFASDIEVTDKATGEKLERT	IRVNHPLT		
	orf110	SX					

40   However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N. gonorrhoeae*:

45	orf110.pep				LLGIASVIGT	LLQONQPQTDYLVKFGSFWA	30
	orf110ng	MSKSRI	SP	TLLSRPWF	FAFFSSMR	FAVALLS	LLGIASVIGT
							VLQONQPQTDYLVKFGPFWT
							60
50	orf110.pep	XIFGFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNVPPFW	REMKSFREKVKEKSLAAMRH			90
	orf110ng	RIFDFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNVPPFW	REMKSFREKVKEKSLAAMRH			120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQ	GFQGTINREDGSVLIAAKK	GTMNKGYIFAHVALIVICL			150
55	orf110ng	SSLLDVKIAPEVAKRYLEV	RGFQGTVSREDGSVLIAAKK	GTMNKGYIXAHVALIVICL			180



-267-

```

orfl10.pep  GGLIDSNLLKLKLGMLTGRIFRTIRRFMPRIKKPESXFGCVQSLIXGQRQYFXRGRVRMWF  210
|  | : | | | | | | | : | | | : | | | | | | | | : | | | | | | | | : | | | |
orfl10ng    GRLINXNLLKLKLGMLAGSIFRNNRRVMPRIKSKPESIWGGVQSLIKGQRQYFQRGKVRMWF  240

5  orfl10.pep  S  211
    |
    orfl10ng   S  241

```

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

```

10      1  MSKSRISPTL  LSRPWFAPFS  SMRFAVALLS  LLGIASVIGT  VLQONQPQTD
      51  YLVKFGPFWT  RIFDFLGLYD  VYASAWFVVI  MMFLVSTSL  CLIRNVPPFW
     101  REMKSFREKV  KEKSLAAMRH  SLLDVKIAP  EVAKRYLEVR  GFQKTVSRE
     151  DGSVLIAAKK  GTMNKWDYIX  AHVALIVICL  GRLINXNLLL  KLGMLAGSIF
     201  RNNRRVMPRI  SKPESIWGGV  QSLIKGQRQY  FQRGKVRMWF  S*

```

15 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

```

20      1  ATGCCGCTCTG  AAACACGCCT  GCCGAACCTT  ATCCGCGTCT  TGATATTTGC
      51  CCTGGGTTTC  ATCTTCCTGA  ACGCCTGTTC  GGAACAAACC  GCGCAAACCG
     101  TTACCTTGCA  AGGCGAAACG  ATGGGCACGA  CCTATACCGT  CAAATACCTT
     151  TCAATAATC  GGGACAAACT  CCCCTCACCT  GCCGAAATAC  AAAAAAGCAT
     201  CGATGACGCG  CTTAAGAAG  TCAACCGGCA  GATGTCCACC  TATCAGCCCG
     251  ACTCCGAAAT  CAGCCGGTTC  AACCAACACA  CAGCCGGCAA  GCCCCCTCCG
     301  ATTTCAAGCG  ACTTCGCACA  CGTTACTGCC  GAAGCCGTCC  GCCTGAACCG
     351  CCTGACACAC  GCGCGCTGG  ACGTAACCGT  CGGCCCTTG  GTCAACCTTT
     401  GGGGATTCGG  CCCCACAAA  TCCGTTACCC  GTGAACCGTC  GCCGGAACAA
     451  ATCAACAGG  CGGCATCTTA  TACGGGCATA  GACAAAATCA  TTTTGAAACA
     501  AGGCAAAGAT  TACGCTTCCT  TGAGCAAAAC  CCACCCCAAG  GCCTATTTGG
     551  ATTTATCTTC  GATTGCCAAA  GGCTTCGGCG  TTGATAAAGT  TGCGGGCGAA
     601  CTGGAAAAAT  ACGGCATTCA  AAATATCTG  GTCGAAATCG  GCGGCGAGTT
     651  GCACGGCAAA  GGCAAAAACG  CGCGCGGCGA  ACCGTGGCGC  ATCGGTATCG
     701  AGCAGCCCAA  TATCGTCCAA  GGCGGCAATA  CGCAGATTAT  CGTCCCGCTG
     751  AACAACCGTT  CGCTTGCCAC  TTCCGGCGAT  TACCGTATTT  TCCACGTCGA
     801  TAAAAACGGC  AAACGCCTCT  CCCATATCAT  CAACCCGAAC  AACAACGAC
     851  CCATCAGCCA  CAACCTCGCC  TCCATCAGCG  TGGTCGCAGA  CAGTGCATG
     901  ACGGCGGACG  GCTTGTCAC  AGGATTATTC  GTATTGGGCG  AAACCGAAGC
     951  CTTAAAGCTG  GCAGAGCGCG  AAAAATCGC  TGTTCCTCTG  ATTGTCAGCG
    1001  ATAAAGGCGG  CTACCGCACC  GCCATGTCTT  CCGAATTTGA  AAAACTGCTC
    1051  CGCTAA

```

This corresponds to the amino acid sequence <SEQ ID 446; ORF111>:

```

45      1  MPSETRLPNF  IRVLIFALGF  IFLNACSEQT  AQTVTLQGET  MGTTYTVKYL
      51  SNNRDKLPSP  AEIQKRIDDA  LKEVNRQMS  YQPDSEISRF  NQHTAGKPLR
     101  ISSDFAHVTA  EAVRLNRLTH  GALDVTVGPL  VNLWGFPGDK  SVTREPSPEQ
     151  IKQAASYTGI  DKIIKQGKD  YASLSKTHPK  AYLDLSSIAK  GFGVDKVAGE
     201  LEKYGIQNYL  VEIGGELHGK  GKNARGEPR  IGIEQPNIVQ  GGNTQIIIVPL
     251  NNRSLATSGD  YRIFHVDKNG  KRLSHIINPN  NKRPISHNLA  SISVADSAM
     301  TADGLSTGLF  VLGETEALKL  AEREKLAFL  IVRDKGGYRT  AMSSEFEKLL
    351  R*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N.*

*meningitidis*:

5	orf111a.pep	10 20 30 40 50 60	MPSETRLPNFIRTLIFALSFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDXLPSP
	orf111	10 20 30 40 50 60	MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSP
10	orf111a.pep	70 80 90 100 110 120	AEIQXRIDDALKEVNRQMSTYQPDSEISRNFNQHTAGKPLRISSDFAHVTAEAVHLNRLTH
	orf111	70 80 90 100 110 120	AEIQXRIDDALKEVNRQMSTYQPDSEISRNFNQHTAGKPLRISSDFAHVTAEAVRLNRLTH
15	orf111a.pep	130 140 150 160 170 180	GALDVTVGPLVNLWGFGDPKSVTREPSPEQIKQAASYTGIDKIILKQGKDYASLSKTHPK
	orf111	130 140 150 160 170 180	GALDVTVGPLVNLWGFGDPKSVTREPSPEQIKQAASYTGIDKIILKQGKDYASLSKTHPK
20	orf111a.pep	190 200 210 220 230 240	AYLDLSSIAGFGVDXVAGELEKYGIQNYLVEIGGELHGKXKNARGEPWRIGIEQPNIVQ
	orf111	190 200 210 220 230 240	AYLDLSSIAGFGVDKVAGELEKYGIQNYLVEIGGELHGKGNARGEPWRIGIEQPNIVQ
25	orf111a.pep	250 260 270 280 290 300	GGNTQIIVPLNNRSXATSGDYRIFHVDKSGKRLSHIINPNNKRPISHNLASISVXADSAM
	orf111	250 260 270 280 290 300	GGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLASISVVADSAM
30	orf111a.pep	310 320 330 340 350	TADGXSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX
	orf111	310 320 330 340 350	TADGLSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCCGCTCTG	AAACACGCCT	GCCGAACCTT	ATCCGCACCT	TGATATTTGC
	51	CCTGAGTTTT	ATCTTCCTGA	ACGCCTGTT	GGAACAAACC	GCGCAAACCG
	101	TTACCCTGCA	AGGTGAAACG	ATGGGCACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACNAACT	CCNCTCACCT	CCCGAAATAC	AAAANCGCAT
	201	CGATGACGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TATCAGCCCG
45	251	ACTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCCTCCGC
	301	ATTTCAAGCG	ACTTCGCACA	CGTTACTGCC	GAAGCCGTCC	ACCTGAACCG
	351	CCTGACACAC	GGCGCGCTGG	ACGTAACCGT	CGGCCCTTGG	GTCAACCTTT
	401	GGGGATTTCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
50	451	ATCAAACAAG	CAGCATCTTA	TACGGGCATA	GACAAAATCA	TTTTGAAACA
	501	AGGCAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAG	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATNANGT	TGCGGGCGAA
	601	CTGGAAAAAT	ACGGCATTCA	AAATATATCT	GTCGAAATCG	GCGGNGAGTT
	651	GCACGGCAAA	GNCAAAAACG	CGCGCGGCGA	ACCTTGGCGC	ATCGGCATCG
	701	AACAGCCCAA	CATCGTCCAA	GGCGGCAATA	CGCAGATTAT	CGTCCCGCTG
55	751	AACAACCGTT	CGNTTGCCAC	TTCCGGCGAT	TACCGTATTT	TCCACGTCGA
	801	TAAAAGCGGC	AAACGCTCT	CCCATATCAT	TAATCCGAAC	AACAACGAC
	851	CCATCAGCCA	CAACCTCGCC	TCCATCAGCG	TGNTCGCAGA	CAGTGCATG
	901	ACGGCGGACG	GCTTNTCCAC	AGGATTATTC	GTATTGGGCG	AAACCGAAGC
	951	CTTAAAGCTG	GCAGAGCGCG	AAAAACTCGC	TGTTTTCCTG	ATTGTCAGGG
60	1001	ATAAAGGCGG	CTACCGCACC	GCCATGTCTT	CCGAATTGTA	AAAACTGCTC
	1051	CGCTAA				

This encodes a protein having amino acid sequence <SEQ ID 448>:

1 MPSETRLPNF IRTLIFALS IFLNACSEQT AQTVTLQGET MGTTYTVKYL  
51 SNNRDXLPSP AEIQXRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR

-269-

101 ISSDFAHVTA EAVHLNRLTH GALDVTVGPL VNLWGFPGDK SVTREPSPEQ  
 151 IKQAASYTGI DKIIKQGKD YASLSKTHPK AYLDLSSIAK GFGVDXVAGE  
 201 LEKYGIQNYL VEIGGELHGK XKNARGEPRW IGIEQPNIVQ GGNTQIIVPL  
 251 NNRSXATSGD YRIFHVDKSG KRLSHIINPN NKRPISHNLA SISVXADSAM  
 301 TADGXSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL  
 351 R\*

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae*:

		10	20	30	40	50	60
	orf111ng	MPSETRLPNLIRALIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLNNRDKLPSP					
	orf111	MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLNNRDKLPSP					
15		10	20	30	40	50	60
	orf111	AKIQKRIDDALKEVNRQMSTYQTDSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH					
20	orf111	AEIQKRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH					
		70	80	90	100	110	120
25	orf111ng	GALDVTVGPLVNLWGFPGDKSVTREPSPEQIKQAASYTGIDKIILOQGDYASLSKTHPK					
	orf111	GALDVTVGPLVNLWGFPGDKSVTREPSPEQIKQAASYTGIDKIILOQGDYASLSKTHPK					
		130	140	150	160	170	180
30	orf111ng	AYLDLSSIAKGFVGVDKVGAGELEKYGIQNYLVEIGGELHGKGNHAGEPWIRIGIEQPNIIQ					
	orf111	AYLDLSSIAKGFVGVDKVGAGELEKYGIQNYLVEIGGELHGKGNARGEPRWIRIGIEQPNIVQ					
		190	200	210	220	230	240
35	orf111ng	GGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLA SISVSDSAM					
	orf111	GGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLA SISVADSAM					
40		250	260	270	280	290	300
	orf111ng	TADGLSTGLFVLGETEALRLAEQEKLAFLIVRDKDGYRTAMSSEFAKLLRX					
	orf111	TADGLSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX					
45		310	320	330	340	350	

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

1 ATGCCGCTCTG AAACACGCCT GCCGAACCTT ATCCGCGCCT TGATATTTGC  
 51 CCTGGGTTTC ATCTTCTCTGA ACGCCTGTTC GGAacaaac CCGCAaaccg  
 101 TTACCCTGCA AGGCGAAAcg aTGGGTACGA CCTATACCGT CAAATACCTT  
 151 TCAATAATC GGGACAAACT CCCCTCCCCT GCCAAAATAC AAAAGCGCAT  
 201 TGATGATGCG CTAAAGAAG TCAACCGGCA GATGTCCACC TACCAGACCG  
 251 ATTCGGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCCTCCGC  
 301 ATTTCAAGCG ATTTGCGACA CGTTACCGCC GAAGCCGTCC GCCTGAACCG  
 351 CCTGACTCAC GCGGCACTGG ACGTAACCGT CGGCCCTTTG GTCACCTTT  
 401 GGGGTTTCGG CCCCACAAA TCCGTTACCC GTGAACCGTC GCCGGAACAA  
 451 ATCAAAACAGG CGGCATCTTA TACGGGCATA GACAAAATCA TTTTGCAACA  
 501 AGGCAAAAGT TACGCTTCCT TGAGCAAAAC CCACCCCAA GCCTATTGG  
 551 ATTTATCTTC GATTGCCAAA GGCTCGGCG TTGATAAAGT TCGGGCGAA  
 601 CTGGAATAAT ACGGCATTCA AAATTATCTG GTCGAAAcg gcggcGAGTT  
 651 GCACGGCAAA GGCAAAATG CGCACGGCGA ACCGTGGCGC ATCGGTATAG  
 701 AGCAACCCAA TATCATCCAA GgcgGCAata CGCAGATTat cgtcccgctg  
 751 aaCaaccggt cgtTGCCAC TTCCGGCGAT TAccgtaTTT tccacgctgA  
 801 TAAAAAcggc aaacgccttt cccacaTCAT CAATCCCaAc aacAAACgac  
 851 ccATCAGcca caacctcgcc tccatcagcg tggctctcAGA CAGTGCAATG  
 901 ACGGCGGACG GTTtatCCAC AGGATTATTT GTTTTAGGCG AAACCGAAGC  
 951 CTTAAGGCTG GCAGAACAAG AAAAActCGC TGTTTTCCTA ATTGTCGGG

1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTTC CAAGCTGCTC  
1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

5 1 MPSETRLPNL IRALIFALGF IFLNACSEQT AQTVTLQGET MGTITYVKYL  
51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMS YQTDSEISRF NQHTAGKPLR  
101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFPGDK SVTREPSPEQ  
151 IKQAASYTGI DKILQQGKD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE  
201 LEKYGIQNYL VEIGGELHGK GKNAGEPWR IGIEQPNIIQ GGNTQIIVPL  
251 NNRLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVSDSAM  
10 301 TADGLSTGLF VLGETEALRL AEQEKLAFL IVRDKDGYRT AMSSEFAKLL  
351 R\*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

sp|P44550|YOJL\_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4  
hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20)  
15 >gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346  
Score = 353 bits (896), Expect = 9e-97  
Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)  
Query: 7 LPNLIRALIFALGFIFLNACSEQT AQTVTLQGETMGTITYVKYLSNNRDKLPSPAKIQKR 66  
20 + LI +I + L AC ++T + ++L G+TMGTTY VKYL + S K +  
Sbjct: 1 MKKLISGIIAVAMALSLAACQKET-KVISLSGKTMGTTYHVKYLDGDSITATSE-KTHEE 58  
Query: 67 IDDALKEVNRQMSYQTDSEISRFNQHT-AGKPLRISDFAHVTA EAVRLNRLTHGALDV 125  
25 I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV  
Sbjct: 59 IEAILKDVNAKMSTYKKDSELSRFNQNTQVNTPIEISADFAKVLAEAIRLNKVTEGALDV 118  
Query: 126 TVGPLVNLWGFPGDKSVTREPSPEQIKQAASYTGIDKIIQQGKDYASLSKTHPKAYLDL 185  
30 TVGP+VNLWGFPG+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL  
Sbjct: 119 TVGFPVNLWGFGEKRPKQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178  
Query: 186 SSIAGFGVDKVAGELEKYGIQNYLVEIGGELHGKGNAGEPWRIGIEQPNIIQGGNTQ 245  
35 SSIAGFGVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +  
Sbjct: 179 SSIAGFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKWPQIAIEKPTTTGERAVE 238  
Query: 246 IIVPLNNRSLATSGDYRI FHVVDKNGKRLSHIINPNKRPI SHNLASISVSDSAMTADGL 305  
35 ++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASI+V++ ++MTADGL  
Sbjct: 239 AVIGLNNMGMASSGDYRIY-FEENGKRFHEIDPKTGYPHQHHLASITVLAPTSMTADGL 297  
Query: 306 STGLFVLGETEALRLAEQEKLAFLIVRDKDGYRTAMSSEFAKL 349  
40 STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL  
Sbjct: 298 STGLFVLGEDKALEVAEKNLAVYLIIRTDNGFVTKSSSAFKKL 341

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 54

45 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 451>:

1 . . CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCAGC CGTCCCCTCA  
51 AAAATTGTGG CTGCGCTTCA TCGGCGGCCG GTCGCATCAA AATATACGGG  
101 GCGGCGCGGC TCGGACGGG TGGCGCAAAG GCGTGCAAT CGGCGCGCAG  
151 GTGTTGTAC GGCAAAATGA AGGCAGCCKA yTGGCAATCG GCGTGATGGG  
50 201 CGGCAGGGCC GCGGACGACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG  
251 gCAGTGATTT GTATGGTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG  
301 TTGCGCGATA AACAAACGGG TgCGTATTG GACGGCTGGT TGCAATACCA  
351 ACGTTTCAAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA  
401 CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG  
55 451 GAAGGCATTG TCGGAAAAGG CAATAATGTG CCGTTTTACC TACAACCGCA  
501 GgCGCAGTTT ACCTACTTGG GCGTAAACGG CCGCTTTACC GACAGCGAGG  
551 GGACGGCGGT CCGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCCGGC  
601 AtTCGGGCAA AAACCCGTTT TGCTTTGCGT AACGGTGTCA ATCTTCAGCC  
651 TTTTGCCGCT TTTAATGTtT TGCACAGGTC AAAATCTTTC GGCGTGGA  
60 701 TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG

This corresponds to the amino acid sequence <SEQ ID 452; ORF35>:

```

1      ..PCRRQGDDVY  AAHASRQKLW  LRFIGGRSHQ  NIRGGAAADG  WRKGVQIGGE
51     VFVRQNEGSX    LATGVMGGRA  GQHASVNGKG  GAAGSDLDYG  GGGVYAAMHW
101    LRDKQTGAYL    DGWLQYQRFK  HRINDENRAE  RYKTKGWATAS  VEGGYNALVA
151    EGIVGKGNNV     RFYLQPPAQF  TYLGVNGGFT  DSEGTAVGLL  GSGQWQSRAG
201    IRAKTRFALR    NGVNLPQFAA  FNVLHRSKSF  GVEMDGKQKT  LAGRTALEGR
251    FGIEAGWKGH    MSA..

```

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

Orf35 5 QGDDVYAAHASRQKLWLRFIGGRSHQNIRGGAA-ADGWRKGQVIGGEVVFVRQNEGSXLAI 63  
 + D++ R+ LWLR I G S+Q ++G A +G+RKGQV+GGEVF QNE + L+I  
 15 virg-h 396 KNSDIFDRTLPRKGLWLRVIDGHSNQWVGKTPAPVEGYRKGQVLGGEVFTWQNESNQLSI 455  
 Orf35 64 GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDKQTGAYLDGWLQYQRFKH 121  
 G+MGG+A Q ++ + ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H  
 virg-h 456 GLMGGQAEQRSTFHNPDNDNLTTGNVKGFGAGVYATWHQLQDKQTGAYADSWMQYQRFH 515  
 20 Orf35 122 RINDENRAERYKTKGWTASVEGGYNALVAEGIVKGKNNVRFYLPQQAQFTYLGVNGGFTD 181  
 RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLPQQAQ TYLGVNG F+D  
 virg-h 516 RINTEDGTERFTSKGITASIEAGYNALLAEHFTKKGNSLRVYLPQQAQLTYLGVNGKFSD 575  
 25 Orf35 182 SEGTAVGLLGSQWQSRAGIRAKTRFALRNGVNLQPFAAFNVLHRSKSFGEVMDGEKQTL 241  
 SE V LLGS Q Q+R G++AK +F+L + ++PFAA N L+ +K FGVEMDGE++ +  
 virg-h 576 SENAHVNLGSRQLQTRVGQAKAQFSLYKNIAIEPFAAVNALYHNKPFGEVMDGERRVI 635  
 Orf35 242 AGRTALEGRFGIEAGWKGHMS 262  
 +TA+E + G+ K H++  
 30 virg-h 636 NNKTAIESOLGVAVKIKSHLT 656

ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N.*

```

35      orf35.pep      10      20      30
                        PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIRG
                        : ||||| ||||| ||||| |||||
    orf35a      310      320      330      340      350      360
      QRLAIPEAEAVLYAQQAYAANTLFLGLRAADRGGDDVYAADPSRQKLWLRFIGGRSHQNIRG

40      orf35.pep      40      50      60      70      80      90
      GAAADGWRKGVQIGGEVFVRQNEGSXLAIGVMGGRAGQHASVNGKGGGAAGSDLYGYGGGV
      ||||| ||||| ||||| ||||| ||||| |||||
    orf35a      370      380      390      400      410      420
      GAAADGRRKGVQIGGEVFVRQNEGSRLAIGVMGGRAGQHASVNGKGGGAAGSYLHGYGGGV

45      orf35.pep      100      110      120      130      140      150
      YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGIV
      ||||| ||||| ||||| ||||| ||||| |||||
    orf35a      430      440      450      460      470      480
      YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGVV

50      orf35.pep      160      170      180      190      200      210
      GKGNNVRFYLPQPAQFTYLGVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRNGVN
      ||||| ||||| ||||| ||||| ||||| |||||
    orf35a      490      500      510      520      530      540
      GKGNNVRFYLPQPAQFTYLGVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRNGVN

55      orf35.pep      220      230      240      250      260
      LQPFAAFNVLHRSKSFVGVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA
      ||||| ||||| ||||| ||||| |||||
    orf35a

```

orf35a LQPF AAFNV LHRSKS FGVEMDGEKQ TLAGRTALEGRFGIEAGWKGHMSARIGYGKRTDGD  
550 560 570 580 590 600

5 orf35a KEAALSLKWLFX  
610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

```

1 ATGTTTCAGAG CTCAGCTTGG TTCAAATACT CGTTCTACCA AAATCGGCGA
51 CGATGCCGAT TTTTCATTTT CAGACAAGCC GAAACCCGGC ACTTCCCATT
101 ATTTTTCAG CCGTAAAACC GATCAAAATT CATCCGAATA TGGGTATGAC
151 GAAATCAATA TCCAAGGTAA AAATACAAT AGCGGCATAC TCGCCGTCGA
201 TAATATGCCC GTTGTAAAGA AATATATTAC AGATACTTAC GGGGATAATT
251 TAAAGGATGC GGTTAAGAAG CAATTACAGG ATTTTATACAA AACAAGACCC
301 GAAGCTTGGG AAGAAAATAA AAAACGGACT GAGGAGGCGT ATATAGAACA
351 GCTTGGACCA AAATTTAGTA TACTCAAACA GAAAAACCCC GATTTAATTA
15 401 ATAAATTGGT AGAAGATTCC GTACTCACTC CTCATAGTAA TACATCACAG
451 ACTAGTCTCA ACAACATCTT CAATAAAAAA TTACACGTCA AAATCGAAAA
501 CAAATCCAC GTCGCGGAC AGGTGTTGGA ACTGACCAAG ATGACGCTGA
551 AAGATTCCCT TTGGGAACCG CGCCGCCATT CCGACATCCA TATGCTGGAA
601 ACTTCGATA ATGCCGCAT CCGCTGAAC ACGAAAGATG AAAAACTGAC
20 651 CGTCCATAAA GCGTATCAGG GCGGTGCGGA TTTCTGTTC GGCTACGACG
701 TGCGGGAGTC GGACAAACCC GCCCTGACCT TTGAAGAAAA AGTCAGCGGA
751 CAATCCGGCG TGGTTTGGGA ACGCCGGCCG GAAATCTGA AAACGCTCGA
801 CGGGCGCAAA CTGATTGCGG CGGAAAAGGC AGACTCTAAT TCGTTGCGT
851 TTAAACAAAA TTACCGGCAG GGACTGTACG AATTATTGCT CAAGCAATGC
25 901 GAAGCGGAT TTTGCTTGGG CGTGACGCGT TTGGCTATCC CCGAGGCGGA
951 AGCGGTTTTA TATGCCCAAC AGGCTTATGC GGCAAATACT TTGTTGGGGC
1001 TGCGTGCCGC CGACAGGGGC GACGACGTGT ATGCCGCCGA TCCGTCCCGT
1051 CAAAAATTGT GGCTGCGCTT CATCGGCGGC CGGTGCGATC AAAATATACG
1101 GGGCGGCGCG GCTGCGGACG GCGGCGCAA AGCGTGCAA ATCGGCGGCG
30 1151 AGGTGTTTGT ACGGCAAAAT GAAGGCAGCC GGCTGGCAAT CGGCGTGATG
1201 GGCGGCAGGG CTGGCCAGCA CGCATCAGTC AACGGCAAAG GCGGTGCGGC
1251 AGGCAGTTAT TTGCATGGTT ATGGCGGGGG TGTATTATGCT GCGTGGCATC
1301 AGTTGCGCGA TAAACAAACG GGTGCGTATT TGGACGGCTG GTTGCAATAC
1351 CAACGTTTCA AACACCGCAT CAATGATGAA AACCGTGCGG AACGCTACAA
35 1401 AACCAAAGGT TGGACGGCTT CTGTGGAAGG CGGCTACAAC GCGCTGTGTG
1451 CGGAAGGCGT TGTGCGAAAA GGCAATAATG TGGGTTTTTA CCTGCAACCG
1501 CAGGCGCAGT TTACCTACTT GGGCGTAAAC GGGCGCTTTA CCGACAGCGA
1551 GGGGACGGCG GTCGGACTGC TCGGCAGCGG TCAGTGGCAA AGCCGCGCCG
1601 GCATTCGGGC AAAAAACCGT TTTGCTTTGC GTAACGGTGT CAATCTTCAG
40 1651 CCTTTTGCCG CTTTTAATGT TTTGCACAGG TCAAAATCTT TCGGCGTGGA
1701 AATGGACGGC GAAAAACAGA CGCTGGCAGG CAGGACGGCG CTCGAAGGGC
1751 GGTTCGGCAT TGAAGCCGGT TGGAAAGGCC ATATGTCCGC ACGCATCGGA
1801 TACGGCAAAA GGACGACGG CGACAAAGAA GCCGCATTGT CGCTCAAATG
1851 GCTGTTTTGA

```

45 This encodes a protein having amino acid sequence <SEQ ID 454>:

```

1 MFRAQLGSNT RSTKIGDDAD FSFSDKPKPG TSHYFSSGKT DQNSSEYGYD
51 EINIQGNYN SGILAVDNMP VVKYITDTY GDNLDKAVK QLQDLYKTRP
101 EAWENKRT EEAYIEQLGP KFSILKQKNP DLINKLVEDS VLTPHSNTSQ
151 TSLNNIFNKK LHVKIENKSH VAGQVLELTK MTLKDSLWEP RRHSDIHMLE
50 201 TSDNARIRLN TKDEKLTVHK AYQGGADFLF GYDVRESDKP ALTFEEKVSG
251 QSGVVLERRP ENLKTLDGRK LIAAEKADSN SFAFKQNYRQ GLYELLKQC
301 EGGFCLGVQR LAIPEAEAVL YAQQAYAANT LFLRLAADRG DDVYAADPSR
351 QKLWLRFIGG RSHQNIIRGA AADGRRKGVQ IGGEVFVRQN EGSRLAIGVM
55 401 GGRAGQHASV NGKGAAGSY LHGYGGVYA AWHQLRDKQT GAYLDGWLQY
451 QRFKHRINDE NRAERYKTKG WTASVEGGYN ALVAEGVVGK GNNVRFYLPQ
501 QAQFTYLVGN GGFTDSEGTA VGLLGSGQWQ SRAGIRAKTR FALRNGVNLQ
551 PFAAFNVLHR SKSFGVEMDG EKQTLAGRTA LEGRFGIEAG WKGHMSARIG
601 YGKRTDGDKE AALSLKWLFX*

```

Homology with a predicted ORF from *N.gonorrhoeae*

60 ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N. gonorrhoeae*:

```

orf35.pep PCRRQGGDDVYAAHASRQKLWLRFIGGRSHQNIIRG 34
:::|:: |::|::|::|::|::|
orf35ngh FTKVQERDDIAIYAQQAQANTLFALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG 370

```

5	orf35.pep	GAA-ADGWRKGVQIGGEVFRQNEGSXLAIGVMGGRAGQHASVNGKG--GAAGSDLYGYG	91
	orf35ngh	KTAPVEGYRKGVQLGGEVFTWQNESNQLSIGLMGGQAEQRSTFRNPDTDNLTGNGVKGF	430
10	orf35.pep	GGVYAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAE	151
	orf35ngh	AGVYATWHQLQDKQTGAYVDSWMQYQRFHRINTEYATERFTSKGITASIEAGYNALLAE	490
15	orf35.pep	GIVGKGNVRFYLPQAQFTYLVNGGGFTDSEGTAVGLLSSGQWQSRAGIRAKTRFALRN	211
	orf35ngh	HFTKKGNSLRVYLPQAQLTYLVNGKFSSENAQVNLLGSRQLQSRVGVQAKAQFAFTN	550
20	orf35.pep	GVNLQPFAAFVNLHRSKSGFVEMDGEKQTLAAGRTALEGRFGIEAGWKGHMSA	263
	orf35ngh	GVTQPFVAVNSIYQKPFGEIDGDRRVINNKTVIETQLGVAARIKSHLTLQASFNRT	610

A partial ORF35ngh nucleotide sequence <SEQ ID 455> is predicted to encode a protein having partial amino acid sequence <SEQ ID 456>:

20	1	..KKLRDRNSEY WKEETYHIKS NGRTPNIPA LFPKHPDFE ENINNSKKIS
	51	FYDKEYTEDY LVGFARGFGV EKRNGEEKP LRQYFKDCVN TENSNDNCK
25	101	ISSFGNYGPI LIKSDIFALA SIKNSHINS EILSVGNIE WLRPTLNKLT
	151	GWQEHLYAGL DPFHYIEVD NSHVIQITD LGALELTNSL WKPRWNSND
30	201	YLITKNAEIR FNTKNESLLV KEDYAGGARF RFAYDLKDKV PEIPVLTFEK
	251	NITGTSDIIF EGKALDNLKH LDGHQIVKVN DTADKDAFRL SSKYRKGIYT
35	301	LSLQQRPEGF FTKVQERDDI AIYAQQQAAA NTLEALRLND KNSDIFDRTL
	351	PRKGLWLRVI DGSNQWVQG KTAPVEGYRK GVQLGGEVFT WQNESNQLSI
40	401	GLMGGQAEQR STFRNPDTDN LTGNGVKFGF AGVYATWHQL QDKQTGAYVD
	451	SWMQYQRFH RINTEYATER FTSKGITASI EAGYNALLAE HFTKKGNSLR
45	501	VYLQPPAQLT YLVNGKFS SENAQVNLLG SRQLQSRVGV QAKAQFAFTN
	551	GVTQPFVAV NSIYQKPFGEIDGDRRVIN NKTVIETQL GVAARIKSHL
	601	TLQASFNRT SKHHHAKQGA LNLQWTF*

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 55

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 457>:

40	1	..GCGGAATATG TTCAGTTCTC TATAGATTG TTCAGTGTGG GTAAATCGGG
	51	GGGCGGTATA CCTAAGGCTA AGCCTGTGTT TGATGCGAAA CCGAGATGGG
45	101	AGGTTGATAG GAAGCTTAAT AAATTGACAA CTCGTGAGCA GGTGGAGAAA
	151	AATGTTACAG AAACGAGAAG AAGGAGTCAG AGTAGTCAGT TTAAAGCCCA
50	201	TGCGCAACGA GAATGGGAAA ATAAAACAGG GTTAGATTTT AATCATTTTA
	251	TAGGTGGTGA TATCAATAAA AAAGGCACAG TAACAGGAGG GCATAGTCTA
55	301	ACCCGTGGTG ATGTACGGGT GATACAACAA ACCTCGGCAC CTGATAAACA
	351	TGGGGT.TTA TCAAGCGACA GTGGAAATTN A

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

45	1	..AEYVQFSIDL FSVGKSGGGI PKAKPVFDAK PRWEVDRLN KLTTRQVEK
	51	NVQETRRRSQ SSQFKAHAQR EWENKTGLDF NHFIGGDINK KGTVTGGHSL
	101	TRGDVRVIQQ TSAPDKHGXL SSDSGNX

Further work revealed further partial nucleotide sequence <SEQ ID 459>:

50	1	..GCAGTGTGCC TnCCGATGCA TGCACACGCC TCAnATTTGG CAAACGATTC
	51	TTTTATCCGG CAGGTTCTCG ACCGTCAGCA TTTCAACCC GACGGGAAAT
55	101	ACCACCTATT CGGCAGCAGG GGGGAAC TTG CCGAGCGCCA GTCTCATATC
	151	GGATTGGGAA AAATACAAAG CCATCAGTTG GGCAACCTGA TGATTCAACA
60	201	GGCGGCCATT AAAGGAAATA TCGGCTACAT TGTCGGCTTT TCCGATCAG
	251	GGCAGGAAGT CCATTCCCCs TTCGACAACC ATGCCTCACA TTCCGATTCT
65	301	GATGAAGCCG GTAGTCCCGT TGACGGATT T AGCCTTTACC GCATCCATTG
	351	GCACGGATAC GAACACCATC CCGCCGACGG CTATGACGGG CCACAGGGCG
	401	GCGGCTATCC CGCTCCCAAA GGCAGGAGG ATATATACAG TTACGACATA

-274-

451 AAAGGCGTTG CCCAAAATAT CCGCCTCAAC CTGACCGACA ACCGCAGCAC  
 501 CGGACAACGG CTTGCCGACC GTTTCACAA TGCCGGTAGT ATGCTGACGC  
 551 AAGGAGTAGG CGACGGATTC AAACGCGCCA CCCGATACAG CCCCGAGCTG  
 601 GACAGATCGG GCAATGCCGC CGAAGCCTTC AACGGCACTG CAGATATCGT  
 651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

1 .AVCLPMHAHA SXLANDSFIR QVLDROHFEP DGKYHLFGSR GELAERQSHI  
 51 GLGKIQSHQL GNLMIQQAII KGNIGYIVRF SDHGHEVHSP FDNHASHSDS  
 101 DEAGSPVDGF SLYRIHWDGY EHHPADGYDG PQGGGYPAKP GARDIYSYDI  
 151 KGVAQNIRLN LTDNRSTGQR LADRFRHNAGS MLTQGVGDGF KRATRYSPDL  
 201 DRSGNAEAF NGTADIVKNI IGAAGEI

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

orf46.pep AEYVQFSIDLFSVSGSGGGIPKAKPVFDAKPRWEVDRLNKLTR 45  
 orf46ng PKTGVPFDGKGFPNFEKHVKYDTKLDIQELSGGGIPKAKPVFDAKPRWEVDRLNKLTR 217  
 20 orf46.pep EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV 105  
 orf46ng EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGAVTGGHSLTRGDV 277  
 orf46.pep RVIQQTSA PDKHXGLSSDSGN 126  
 25 orf46ng RVIQQTSA PDKHXGLSSDSGN 298

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

1 .RLKHCCHAR LGSFHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKPKQC  
 30 51 RTRHRSRQQY LYGSHPHQRD WSCPGKIQLG RHHGTSCRAV ADXRDRICER  
 101 EIRRQXQXCR CRLGKIPSL IPKYPLKLEQ RYKENITSS TVPPSNGKNV  
 151 KLADQRHPKT GVPFDGKGFP NFEKHVKYDT KLDIQELSGG GIPKAKPVFD  
 201 AKPRWEVDRL LNKLTTRQV EKNVQETRRR SQSSQFKAHA QREWENKTGL  
 251 DFNHFIGGDI NKKGAVTGGH SLTRGDVIRVI QQTSA PDKHG VLSSDSGN\*

35 Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

1 TTGGGCATTT CCCGCAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG  
 51 CCTGCCGATG CATGCACACG CCTCAGATTT GGcaAACGAT CCCTTTATCC  
 101 GgCaggttct CGaccGTCAG CATTTCGaac ccgacggGaa ATACCaCCTA  
 151 TTcggCaGCA GGGGGGAGCT TgccnagcGC aacggccATa tcggattggG  
 201 aaacaTAcAa Agccatcagt tGggccacct gatgattcaa caggcggccg  
 251 ttgaaggaaA TAtcgGctac attgtccgct tttccgatca cgggcacaaa  
 301 ttccattcgc ccttcGAcAa ccaTGCCTCA CATTCCGATT CTGACGAAGC  
 351 CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGGACGGAT  
 401 ACGAACACCA TCCCGCCGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT  
 451 CCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAAGGCGT  
 501 TGCCCAAAAT ATCCGCTCA ACCTGACCGA CAACCGCAGC ACCGGACAAC  
 551 GGCTTGCCGA CCGTTTCCAC AATGCCGCGC CTATGCTGAC GCAAGGAGTA  
 601 GGGCAGCGAT TCAAACGCGC CACCCGATAC AGCCCCGAGC TGGACAGATC  
 651 GGGCAATGcc gccGAAGCCT TCAACGGCAC TGCAGATATC GTCAAAAACA  
 701 TCATCGGCGC GGCAGGAGAA ATTGTCGGCG CAGGCGATGC CGTGCagGGT  
 751 ATAAGCGAAG GCTCAAACAT TGCTGTCATG CACGGCTTGG GTCTGCTTTC  
 801 CACCGAAAAC AAGATGGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC  
 851 TCAAAGACTA TGCCGAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC  
 901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA TGGCAGCCAT  
 951 CCCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCG  
 1001 TCACGGCACA TCCTGTCAAG CGGTCGCAGA TGGGCGCGAT CGCATTGCCG  
 1051 AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA  
 1101 ATACCCGTCC CTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC



-275-

1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC  
 1201 AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT  
 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA  
 1301 AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTAA GGCTAAGCCT  
 1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT  
 1401 GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAACG AGAAGAAGGA  
 1451 GTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAAAAATAA  
 1501 ACAGGGTTAG ATTTTAAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG  
 1551 CACAGTAACA GGAGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC  
 1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGAA  
 1651 ATTAATAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA  
 1701 AGTGATGACC AAGCACACCA TGTTCCCAA AGATTGGGAT GAGGCTAGAA  
 1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT  
 1801 AATAAATGGC AGGGTACAAG TAAATCGGGT ATTAATAATG AAGGATTTAC  
 1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

1 LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDRO HFEPDGKYHL  
 51 FGSRGELAXR NGHIGLGNIQ SHQLGHLMIQ QAAVEGNIGY IVRFSDHGKH  
 101 FHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHPAD GYDGPQGGY  
 151 PAPKGARDIY SYDIKVAQN IRLNLTNRS TGQRLADRFH NAGAMLTQGV  
 201 GDGFKRATRY SPELDRSGNA AEAFTGTADI VKNIIGAAGE IVGAGDAVQG  
 251 ISEGSNIAMV HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP  
 301 NAAQIEAVS NIFMAAIPK GIGAVRGKYG LGGITAHFVK RSQMGAIALP  
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYKENI TSSTVPPSNG  
 401 KNVKLADQRH PKTGVFPDGG GF'PNFEKHVK YDTKLDIQEL SGGGIPKAKP  
 451 VFDAPRWEV DRKLNKLTR EQVEKNVQET RRRSQSSQFK AHQREWENK  
 501 TGLDFNHFIG GDINKKGTVT GGHSLTRGDV RVIQQTSA PD KHGVYQATVE  
 551 IKKPDGSWEV KTKKGKVMK KHTMFPKDWD EARIRAEVTS AWESRIMLKD  
 601 NKWQTSKSG IKIEGFTEPN RTAYPIYE\*

ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

		10	20	30	40		
orf46-1.pep		AVCLPMHAHASXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER					
orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR						
		10	20	30	40	50	60
orf46-1.pep		50	60	70	80	90	100
orf46ng-1	QSHIGLGKIQSHQLGNLMIQQAIAKGNIGYIVRFSHDGHEVHSPFDNHASHSDSDEAGSP						
	NGHIGLGNIQSHQLGHLMIQQAIAVEGNIGYIVRFSHDGKHKTHSPFDNHASHSDSDEAGSP						
		70	80	90	100	110	120
orf46-1.pep		110	120	130	140	150	160
orf46ng-1	VDGFSLYRIHWDGYEHPADGYDGPQGGGYPAPKGARDIYSYDIKVAQNIRLNLTNRS						
	VDGFSLYRIHWDGYEHPADGYDGPQGGGYPAPKGARDIYSYDIKVAQNIRLNLTNRS						
		130	140	150	160	170	180
orf46-1.pep		170	180	190	200	210	220
orf46ng-1	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDERSGNAEAFNGTADIVKNIIGAAGE						
	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDERSGNAEAFNGTADIVKNIIGAAGE						
		190	200	210	220	230	240
orf46-1.pep	I						
orf46ng-1	IVGAGDAVQGISSEGSNIAMVHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP						
		250	260	270	280	290	300

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of

*N. meningitidis*:

10 20 30 40 50 60

-276-

5	orf46a.pep	LGISRKISLILSLAVCLPMHAHASDLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER
	orf46ng-1	LGISRKISLILSLAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR
10	orf46a.pep	SGHIGLGNIQSHQLGNLFIQQAAIKGNIGYIVRFSHDHGHEVHSPFDNHASHSDSDEAGSP
	orf46ng-1	NGHIGLGNIQSHQLGHLMIQQAAVEGNIGYIVRFSHDHGHEVHSPFDNHASHSDSDEAGSP
15	orf46a.pep	VDGFSLYRIHWDGYEHHPADGYDGPQGGGYPAPKGARDIYSYDIKVAQNIRLNLTDNRS
	orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGGYPAPKGARDIYSYDIKVAQNIRLNLTDNRS
20	orf46a.pep	TGQRLVDRFHNTGSMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE
	orf46ng-1	TGQRLADRFHNAAGAMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE
25	orf46a.pep	IVGAGDAVQGISSEGSNIAVMHGLGLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
	orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
30	orf46a.pep	NAAQGIEAVSNIFTAVIPVKIGIAVRGKYGLGGITAHPVKRSQMGEIALPKGKSAVSDNF
	orf46ng-1	NAAQGIEAVSNIFMAAIPKIGIAVRGKYGLGGITAHPVKRSQMGAIALPKGKSAVSDNF
35	orf46a.pep	ADAAYAKYPSPYHSRNIERNLEQRYGKENITSSTVPPSNGKNVLANRHPKTKVPFDGK
	orf46ng-1	ADAAYAKYPSPYHSRNIERNLEQRYGKENITSSTVPPSNGKNVLANRHPKTKVPFDGK
40	orf46a.pep	GFPNFEKDVKYDTRINTAVPQVN----PIDEPVFN--PKGSVGSASHSWSITARIQYAKLP
	orf46ng-1	GFPNFEKHVKYDTKLD--IQELSGGGIPKAKPVFDAKPRWEVDKRLN-KLTTRREQVEKNV
45	orf46a.pep	RQGRIRYIPKKNYSAPSAPLPKGPNNGYLDKFGNEWTGKPSRTKGQFEWDVQLSKTGREQ
	orf46ng-1	QETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDVRVIQOTS

The complete length ORF46a DNA sequence &lt;SEQ ID 465&gt; is:

55	1	TTGGGCATTT	CCCGCAAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CATGCACACG	CCTCAGATTT	GGCAAACGAT	TCTTTTATCC
60	101	GGCAGGTCT	CGACCGTCAG	CATTTTCAAC	CCGACGGGAA	ATACCACCTA
	151	TTCGGCAGCA	GGGGGGAAC	TGCCGAGCGC	AGCGGTCATA	TCGGATTGGG
65	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTTCATCCAG	CAGGCGGCCA
	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCCGATCA	CGGGCACGAA
70	301	GTCCATTCCC	CCTTCGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTGACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
75	401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGTAT
	451	CCCCTCCCA	AAGGCGCGAG	GGATATATAC	AGCTACGACA	TAAAAGGCGT
80	501	TGCCCCAAAT	ATCCGCCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC
	551	GGCTTGTCGA	CCGTTTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
85	601	GGCGACGGAT	TCAAACGCGC	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCGAAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
90	701	TCATCGGCGC	GGCAGGAGAA	ATTGTCCGCG	CAGGCGATGC	CGTGCAGGGT
	751	ATAAGCGAAG	GCTCAAACAT	TGCTGTTATG	CACGGCTTGG	GTCTGCTTTC
95	801	CACCGAAAAC	AAGATGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC

5 851 TC A A G A C T A T G C C G C A G C A G C C A T C C G C G A T T G G G C A G T C C A A A A C C C C  
 901 A A T G C C G C A C A A G G C A T A G A A G C C G T C A G C A A T A T C T T T A C G G C A G T C A T  
 951 C C C C G T C A A A G G A T T G G A G C T G T T C G G G G A A A A T A C G G C T T G G G C G G C A  
 10 1001 T C A C G G C A C A T C C T G T C A A G C G G T C G C A G A T G G G C G A G A T C G C A T T G C C G  
 1051 A A A G G G A A A T C C G C C G T C A G C G A C A A T T T T G C C G A T G C G G C A T A C G C C A A  
 1101 A T A C C C G T C C C C T T A C C A T T C C C G A A A T A T C C G T T C A A A C T T G G A G C A G C  
 1151 G T T A C G G C A A A G A A A A C A T C A C C T C C T C A A C C G T G C C G C C G T C A A A C G G A  
 1201 A A G A A T G T G A A A C T G G C A A A C A A C G C C A C C G A A G A C C A A A G T G C C G T T  
 1251 T G A C G G T A A A G G G T T T C C G A A T T T T G A A A A A G A C G T A A A A T A C G A T A C G A  
 1301 G A A T T A A T A C C G C T G T A C C A C A A G T G A A T C C T A T A G A T G A A C C C G T C T T T  
 1351 A A T C C T A A A G G T T C T G T C G G A T C G G C T C A T T C T T G G T C T A T A A C T G C C A G  
 1401 A A T T C A A T A C G C A A A A T T A C A A G G C A A G G T A G A A T C A G A T A T A T C C C A C  
 1451 C T A A A A T T A C T C T C C T T C A G C A C C G T A C C A A A G G A C C T A A T A A T G G A  
 1501 T A T T T G G A T A A T T T G G T A A T G A A T G G A C T A A A G G T C C A T C A A G A A C T A A  
 1551 A G G T C A A G A A T T T G A A T G G G A T G T T C A A T T G T C T A A A A C A G G A A G A G A G C  
 1601 A A C T T G G A T G G G C T A G T A G G A T G G T A A G C A T T T A A A T A T A T C A A T T G A T  
 1651 G G A A G A T T A C A C A A A A T G A

This corresponds to the amino acid sequence <SEQ ID 466>:

20 1 L G I S R K I S L I L S I L A V C L P M H A H A S D L A N D S F I R Q V L D R Q H F E P D G K Y H L  
 51 F G S R G E L A E R S G H I G L N I Q S H Q L G N L F I Q Q A A I K N I G Y I V R F S D H G H E  
 101 V H S P F D N H A S H S D S D E A G S P V D G F S L Y R I H W D G Y E H H P A D G Y D G P Q G G G Y  
 151 P A P K G A R D I Y S Y D I K V A Q N I R L N L D N R S T G Q R L V D R F H N T G S M L T Q G V  
 201 G D G F K R A T R Y S P E L D R S G N A A E A F N G T A D I V K N I I G A A G E I V G A G D A V Q G  
 25 251 I S E G S N I A V M H G L G L L S T E N K M A R I N D L A D M A Q L K D Y A A A A I R D W A V Q N P  
 301 N A A Q G I E A V S N I F T A V I P V K G I G A V R G K Y G L G G I T A H P V K R S Q M G E I A L P  
 351 K G K S A V S D N F A D A A Y A K Y P S P Y H S R N I R S N L E Q R Y G K E N I T S S T V P P S N G  
 401 K N V K L A N K R H P K T K V P F D G K G E P N F E K D V K Y D T R I N T A V P Q V N P I D E P V F  
 451 N P K G S V G S A H S W S I T A R I Q Y A K L P R Q G R I R Y I P P K N Y S P S A P L P K G P N N G  
 501 Y L D K F G N E W T K G P S R T K Q E F E W D V Q L S K T G R E Q L G W A S R D G K H L N I S I D  
 30 551 G K I T H K \*

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 35 Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

40 1 A T G A A T A T T C A C A C C T G C T C T C C A A A C A A T G G A C G T G C C G C C A T T C C T  
 51 G C C G A A A C G G C T G C T G C T G T C C C T G C T G A T A C T G C T T G C C C C A A T G C G G  
 101 T G T T T T G G G T T T T G G C A C T G C T G A C C G C C A C G C C C G C C C G A T T G T C A A T  
 151 T T G G A C T A T C T T C C C G C C G C G C T G C T G A T C G C C C T G C C T T G C G T T C G T  
 201 C A A A A T T G C C G G C G T A T T G G C G T T T T G G C T G G C G G T T T T G T T T G A C G G G C  
 251 T G A T G A T G G T G A T C C A A C T C T T C C C T T T T A T G G A T C T C A T C G G C G C C A T C  
 301 A A C C T C G T C C C C T T C A T C C T G A C C G C C C C G C C C T T A T C A G A T A A T G A C  
 351 C G G G C T G . . .

45 This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

1 M N I H T L L S K Q W T L P P F L P K R L L S L L I L L A P N A V F W V L A L L T A T A R P I V N  
 51 L D Y L P A A L L I A L P W R F V K I A G V L A F W L A V L F D G L M M V I Q L F P F M D L I G A I  
 101 N L V P F I L T A P A P Y Q I M T G L . . .

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

50 1 A T G A A T A T T C A C A C C T G C T C T C C A A A C A A T G G A C G T G C C G C C A T T C C T  
 51 G C C G A A A C G G C T G C T G C T G T C C C T G C T G A T A C T G C T T G C C C C A A T G C G G  
 101 T G T T T T G G G T T T T G G C A C T G C T G A C C G C C A C G C C C G C C C G A T T G T C A A T  
 151 T T G G A C T A T C T T C C C G C C G C G C T G C T G A T C G C C C T G C C T T G C G T T T C G T  
 201 C A A A A T T G C C G G C G T A T T G G C G T T T T G G C T G G C G G T T T T G T T T G A C G G G C  
 55 251 T G A T G A T G G T G A T C C A A C T C T T C C C T T T T A T G G A T C T C A T C G G C G C C A T C  
 301 A A C C T C G T C C C C T T C A T C C T G A C C G C C C C G C C C T T A T C A G A T A A T G A C  
 351 C G G G C T G T T G C T G T A T A T G C T G G C G A T G C C G T T T G T G T T G C A G A A A G

-278-

5  
10  
15  
20

```

401 CCGCGGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCCGTTGTG
451 GCGGCAGCCG GCTATTTTAC CGGCCATTG AGTTACTACG ACCGGGGTCG
501 GATGGCCAAT ATCTTCGGCG CAAACAACCT CTACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTG GCCGAATCTT
701 GGGGGCTGCC GGCCAAATCCG GAACTTCAAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC
901 TGCCTCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
951 CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACTTGC
1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTGCGCG AAGTGTGCGC
1101 ATTTTTCAAA AAACACGACA AGGGACTGTT TTACTGGATG ACGCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCCGACATT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCGCGC CGAAACCGAC CTCTGCCGCA ATTTCCAGCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
1351 AACCTCAATG AAACCTTCG CTACCTCAA CAGGGGCACG TCGCCTGGCT
1401 GAACTTCAAA ATCAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

25  
30

```

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFVWLAL LTATARPIVN
51 LDYLPALLI ALPWRVFKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAVCAAVV
151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFTITAGL
201 VDPVFLPLGN QORAATHLNE PKSQKILFIV AESWGLPANP ELQNAITFAKL
251 LAQKDRFSVW ESGSFFFIGA TVEGEMRELC AYGLRGLFAL RRAPDEKFAF
301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKKTG
351 AIFGVCDSSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHLRLK
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVPV
451 NLNETFRYLYK QGHVAWLNFK IK*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N. meningitidis*:

40  
45  
50

```

          10      20      30      40      50      60
orf48.pep  MNIHTLLSKQWTLPPFLPKRLLLSLILLA PNAVFVWLAL LTATARPIVNLDYLPALLI
          |||||
orf48a     MNIHTLLSKQWTLPPFLPKRLLLSLILLA PNAVFVWLAL LTATARPIVNLXYPALLI
          10      20      30      40      50      60

          70      80      90      100     110     119
orf48.pep  ALPWRVFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL
          |||||
orf48a     ALPWRXVKIXGVLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVPFXTAPALYQIMTGLL
          70      80      90      100     110     120

orf48a     LLYMLAMPFVLQKAAKTDFRHIAACA VVAAGYFTGHLSXYDRGRMANIFGANNFYA
          130     140     150     160     170     180

```

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

55  
60

```

1 ATGAATATTC ACACCTGCT CTCAAACAA TGGACGCTGC CGCCATTCCT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTNNCC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGANTACC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GCGGTNTCGT
201 CAAAAATGNC GCGGTATTGG CGTNTTGGCT GCGGTTTGTG TTTGACGGGC
251 TGATGATGGT GATCCAACTC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCNT GACCGCCCCC GCCCTTTATC AGATAATGAC
351 CGGGCTGTTA CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
401 CCGCCGCCAA AACCGACTTC CGACACATTG CCGCCTGTGC CGCCGTTGTG
451 GTGGCAGCCG GCTATTTTAC CGGCCATTG AGTTANTACG ACCGGGGGCG

```

-279-

501 GATGGCCAAT ATCTTCGGCG CAAACAACCTT CTATTACGCC AAAAGTCAGG  
 551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG  
 601 GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA  
 5 651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT  
 701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG  
 751 CTGGCGCAAA AAGANCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT  
 801 CATCGGCGCG ACGATCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG  
 851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC  
 901 TGCTCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA  
 10 951 CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG  
 1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC  
 1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTGCGCG AAGTGTGCGC  
 1101 ANTTTTTCAA AAACACGACA AGGGACTGTT TTAAGTGATG ACGCTGACCA  
 1151 GCCACGCCGA CTATCCCGAA TCNGACATTT TCAACACAG GCTCAAATGC  
 15 1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC NTCTGCCGCA ATTTCAGCCT  
 1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCGGAAA  
 1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC  
 1351 AACCTCAATG AAACCTTCCG CTACCTCAA CAGGGGCACG TCGNCTGGCT  
 1401 GAACTTCAA ATCAAATA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNIHTLLSKQ WTLPPFLPKR LLSLLILLX PNAVFWLAL LTATARPIVN  
 51 LXLYPAALLI ALPWRXVKIX GVLAXWLAVL FDGLMMVIQL FPFMDLIGAI  
 101 NLVPFXTAP ALYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAACAADV  
 151 VAAGYFTGHL SXYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL  
 201 VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQNATFAKL  
 25 251 LAQKXRFSVW ESGSFPIGA TIEGEMRELC AYGGLRGFAL RRAPDEKFAR  
 301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQEIKT AENLIGKKT  
 351 AIFGGVCDSE LFGEVSAXFK KHDKGLFYWM TLTSHADYPE SDIFNHLKLC  
 401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPFVG  
 30 451 NLNETFRYLK QGHVXWLNFK IK\*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

		10	20	30	40	50	60
orf48a.pep		MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNVFWVLALITATARPIVNLXLYPAALLI					
35	orf48-1	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALITATARPIVNLXLYPAALLI					
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf48a.pep	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVPFXTAPALYQIMTGLL					
	orf48-1	ALPWRXVKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFXTAPALYQIMTGLL					
		70	80	90	100	110	120
		130	140	150	160	170	180
45	orf48a.pep	LLYMLAMPFVLQKAAKTDFRHIAACAADVVAAGYFTGHLXSYDRGRMANIFGANNFYA					
	orf48-1	LLYMLAMPFVLQKAAKTDFRHIAACAADVVAAGYFTGHLXSYDRGRMANIFGANNFYA					
		130	140	150	160	170	180
50	orf48a.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLPANP					
	orf48-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLPANP					
		190	200	210	220	230	240
55	orf48a.pep	ELQNATFAKLLAQKXRFSVWESGSFPIGATIEGEMRELCAYGGLRGFALRRAPDEKFAR					
	orf48-1	ELQNATFAKLLAQKXRFSVWESGSFPIGATIEGEMRELCAYGGLRGFALRRAPDEKFAR					
		250	260	270	280	290	300
60	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKAENLIGKKTCAIFGGVCDSE					
	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKAENLIGKKTCAIFGGVCDSE					
		310	320	330	340	350	360
65	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKAENLIGKKTCAIFGGVCDSE					
	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKAENLIGKKTCAIFGGVCDSE					
		310	320	330	340	350	360

-280-

		370	380	390	400	410	420
	orf48a.pep	LFGEVSAXFKKHDKGLFYWMTLTSHADYPESDIFNHLRKCTEYGLPAETDXCRNFSLHTQ					
5	orf48-1	LFGEVSAXFKKHDKGLFYWMTLTSHADYPESDIFNHLRKCTEYGLPAETDLCRNFSLHTQ					
		370	380	390	400	410	420
	orf48a.pep	FFDQLADLIQRPEMKGTEVIIIVGDHPPVGNLNETFRYLKQGHVXWLNFKIKX					
10	orf48-1	FFDQLADLIQRPEMKGTEVIIIVGDHPPVGNLNETFRYLKQGHVAWLNFKIKX					
		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*

15 *gonorrhoeae*:

	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLDYLPAAALLI	60
		:	
	orf48ng	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLDYLPAAALLI	60
20	orf48.pep	ALPWRFVKIAGVLAFWLAVLFDGLMMVIQLFFPMDLIGAINLVPFILTAPAPYQIMTGL	119
	orf48ng	ALPWRFVKIAGVLAFWPAVLFDGLMMVIQLFFPMDLIGAINLVPFILTAPAPYQIMTGLL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

25	1	MNIHALLSEQ	WTLPPFLPKR	LLLSLLILLA	PNAVFWVLAL	LTATARPIVN
	51	LDYLPAAALLI	ALPWRFVKIA	GVLAFWPAVL	FDGLMMVIQL	FFPMDLIGAI
	101	NLVPFILTAP	APYQIMTGLL	LLYMLAMPFV	LQKAAVKTD	RHIAVCAAVV
	151	AAARYFTGP	ELLRTGGRWQ	YVQHRRLLS	GSRASFRRRQ	KADVLRLRGN
	201	PYASMGNGG	..			

## 30 Further work identified the complete gonococcal DNA sequence &lt;SEQ ID 475&gt;:

	1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTCCT
	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATGCGG
	101	TGTTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCCGCCC	GATTGTCAAT
	151	TTGGACTACC	TCCCGCGCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTTTCGT
35	201	CAAAATTGCC	GGCGTATTGG	CGTTTTGGCC	GGCGGTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAACTC	TCCCTTTTA	TGGACCTCAT	CGGCGCATC
	301	AACCTCGTCC	CCTTCATCCT	GACCGCCCCC	GCCCCCTATC	AGATAATGAC
	351	CGGGCTGTTG	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAAAAAG
	401	CCGCCGTCAA	AACCGACTTC	CGACACATTG	CCGTCTGTGC	CGCCGTTGTG
40	451	GCGGCAGCCG	GCTATTTCAC	CGGCCATTTG	AGTTACTACG	ACCGGGGCGC
	501	GATGGCCAAT	ATCTTCGGCG	CAACAACCTT	CTATTACGCC	aAAAGTCAGG
	551	CGATGCTCTA	CACCGTCAGC	CAGAATGCCG	ACTTTATTAC	CGCCGgcctG
	601	GTCGACCCCG	TCTTCCTCCC	CTTGGGCAAT	CAGCAGCGTG	CCGCCACGCG
	651	GCTGAGTGAG	CCGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCCGAATCTT
45	701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAAA	ACGCCACTTT	TGCCAAACTG
	751	CTGGCGCAAA	AAGACCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCCTT
	801	CATCGGCGCG	ACGGTCGAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
	851	GTTTGCGCGG	GTTTCGCACTG	CGCCGCGCGC	CCGACGAAAA	ATTTGCCCGC
	901	TGCCTCCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TGCGGATGCA
50	951	CGGCGCGGGT	AGTTTCGCTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGCGGG
	1001	GCTTTCAAAA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGGTGT	CGACAGCGAG	CTGTTTCGGC	AAGTGTCCGC
	1101	ATTTTCAAAA	AAACACGACA	AGGGACTGTT	TACTGGATG	ACGCTGACCA
	1151	GCCACGCGGA	CTATCCCGAA	TCCGACATTT	TCAACCACAG	GCTCAAATGC
55	1201	ACGAATACG	GCCTGCCCGC	CGAAACCGAC	CTCTGCCGCA	ATTTACGCCT
	1251	GCACACCCAA	TtcttcgACC	AACTGGCGGA	TTTGATCCGA	CGCCCCGAAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	GCCCGTCGGC
	1351	AACCTCAATG	AAACCTCCG	CTACCTCAAA	CAGGGACACG	TGCCTGGCT
	1401	GCATTCAAAA	ATCAAATAA			

## 60 This encodes a protein having amino acid sequence &lt;SEQ ID 476; ORF48ng-1&gt;:

-281-

1 MNIHALLSEQ WTLPPFLPKR LLLSILLILLA PNAVFWVLAL LTATARPIVN  
 51 LDYLPALLI ALPWRVFKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI  
 101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD F RHIAVCAAVV  
 151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL  
 201 VDPVFLPLGN QQRAATRLSE PKSOKILFTV AESWGLPGNP ELQNATFAKL  
 251 LAQKDRFSVW ESGSFFFIGA TVEGEMRELC AYGGLRGFAL RRAPDEK FAR  
 301 CLPNRLKQEG YATFAMHGAG SSYDRFSWY PRAGFQIKT AENLIGKKT C  
 351 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHR LKC  
 401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPPV G  
 10 451 NLNETFRYLK QGHVAWLHFK IK\*

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

		10	20	30	40	50	60
	orf48-1.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNDYLP	AALLI				
15	orf48ng-1	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNDYLP	AALLI				
		10	20	30	40	50	60
	orf48-1.pep	ALPWRVFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVVPFILTAP	APYQIMTGLL				
20	orf48ng-1	ALPWRVFKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVVPFILTAP	APYQIMTGLL				
		70	80	90	100	110	120
	orf48-1.pep	LLYMLAMPFVLQKAAAKTDFRHI	AVCAAVVAAAGYFTGHLSYDRGRMANIFGANNFYA				
25	orf48ng-1	LLYMLAMPFVLQKAAAKTDFRHI	AVCAAVVAAAGYFTGHLSYDRGRMANIFGANNFYA				
		130	140	150	160	170	180
	orf48-1.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSOKILFIVA	ESWGLPANP				
30	orf48ng-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATRLSEPKSOKILFIVA	ESWGLPGNP				
		190	200	210	220	230	240
	orf48-1.pep	ELQNATFAKLLAQKDRFSVWESGSFPPFIGATVEGEMRELCAYGGLRGFAL	RRAPDEK FAR				
35	orf48ng-1	ELQNATFAKLLAQKDRFSVWESGSFPPFIGATVEGEMRELCAYGGLRGFAL	RRAPDEK FAR				
		250	260	270	280	290	300
	orf48-1.pep	CLPNRLKQEGYATFAMHGAGSSYDRFSWYPRAGFQIKTAENLIGKKTCAI	FGGVCDSE				
40	orf48ng-1	CLPNRLKQEGYATFAMHGAGSSYDRFSWYPRAGFQIKTAENLIGKKTCAI	FGGVCDSE				
		310	320	330	340	350	360
	orf48-1.pep	LFGEVSAFFKKHDKGLFYWM	TLTSHADYPESDIFNHR LKCTEYGLPAETDLCRNFSLHTQ				
45	orf48ng-1	LFGEVSAFFKKHDKGLFYWM	TLTSHADYPESDIFNHR LKCTEYGLPAETDLCRNFSLHTQ				
		370	380	390	400	410	420
	orf48-1.pep	FFDQLADLIQRPEMKGTEVIIVGDHPPPVGNLNETFRYLKQGHVAWL	NFKIKX				
50	orf48ng-1	FFDQLADLIRPEMKGTEVIIVGDHPPPVGNLNETFRYLKQGHVAWL	NFKIKX				
		430	440	450	460	470	
	orf48-1.pep	FFDQLADLIQRPEMKGTEVIIVGDHPPPVGNLNETFRYLKQGHVAWL	NFKIKX				
55	orf48ng-1	FFDQLADLIRPEMKGTEVIIVGDHPPPVGNLNETFRYLKQGHVAWL	NFKIKX				

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 57**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

      1  ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCACAAA TCATCATCGT
    51  TACTTTGAGT ATCGCCACGC TTGCCGCGC CGGCATCGCT ATGTCGCGCG
15  101  GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
    151  GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCC CGCCGATTGA
    201  AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
    251  CTTCCGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
    301  AGTGCGGTTT TGGCTTTGGT TTTCTTGCA CTGGGCGC.G TAGCGCCGAA
10  351  CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAATAT AACGGGCAAT
    401  TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

      1  ..VSGRYRALDR VSKIIIVTLS IATLAAAGIA MSRCMQMSD FIEPTPWTLA
    51  GLGFLIALMG WMPAPIEISA INSLWVTEKQ RINPSEYRDG IFEFNVGYIA
15  101  SAVLALVFLA LGXVAPNGNG XTVQMAGGKY NGQLINMYA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

      1  ATGTCGGAAC AACATATTTT GACTTGGAAA AGTAAAATCA ACGCATTGGG
    51  TCCGGGGATC ATGATGGCTT CGGCGGCGGT CGGCGGTTTC CACCTGATTG
15  101  CCTCGACGCA GCGGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
    151  ATCCTGACCA ACCTCTTCAA ATACCGGTTT TTCCGCTTCA GCGCGCATTG
    201  CACGCTGGAC ACGGCAAGA GCCTGATTGA AGGTATGCC GAGAAAAGCC
    251  GCGTTATTTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
    301  AACGCGGGCG CGGTGCGCAT TGTAAACGCC GCCATCGTCA AAATGGCGAT
25  351  TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
    401  CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
    451  TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCGCGCGG
    501  CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
    551  CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
30  601  ATGCCGCGCG CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
    651  AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTCG
    701  ACGTCGGTTA TATCGCCAGT GCGGTTTGGT CTTTGGTTTT CCTTGCATCG
    751  GCGCGGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
    801  CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTTACC ATCGGCGGCT
35  851  GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCAGC
    901  ACGATTACCG TCGTGGACGG CTATGCCCGT GCCATTGCCG AACCCTGCGG
    951  CCTGCTGCGC GGAAAAGACA AAACGGGCAA CGCCGAATTC TTTGCCTGGA
100 1001  ATATTGGGGT GCGGGGCGAG GGTTTGGCGG TGATTTTCTG GTTTGACGGC
    1051  GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
15  1101  CCCTGTGTTT GCCTGGCTGA ATTACCGTTT GGTAAAGGT GATGAAAAC
40  1151  ACAAACACAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTAT
    1201  CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
    1251  ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

      1  MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASTQAGA LYGWQIALII
45  51  ILTNLFKYPP FRFSAHYTLD TGKSLIEGYA EKSRYVLWVF LILCILSATI
    101  NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
    151  SKIIIVTLSI ATLAAAGIAM SRGMQMSDF IEPTPWTLAG LGFLIALMGW
    201  MPAPIEISAI NSLWVTEKQR INPSEYRDI FDFNVGYIAS AVLALVFLAL
50  251  GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
    301  TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIVVAGS GLAVIFWFDG
    351  VMANLLKFAM IAAFSAPVF AWLNRYRLVG DEKHKLTSGM NALALAGLIY
    401  LTGFTVLELL NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N.meningitidis*:



-283-

					10	20	30
	orf53.pep				VSGRYRALDRVSKI	IIIVTL	SIATLAAAGIA
5	orf53a	AAIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKI	IIIVTL	SIATLAAAGIA			
		110	120	130	140	150	160
			40	50	60	70	80
	orf53.pep	MSRGMQMOSDFIEPTPWTL	LAGLGLIALMGWMPAP	IEISAINSLWVTEKQ	RINPSEYRDG		
10	orf53a	MSRGMQMOSDFIEPTPWTL	LAGLGLIALMGWMPAP	IEISAINSLWVTEKQ	RINPSEYRDG		
		170	180	190	200	210	220
			100	110	120	130	139
	orf53.pep	IFEFNVG	YIASAVLALVFLALGX	VAPNGNGXTVQ	MAGGKYNGQLINMYA		
15	orf53a	IFDFNVGYIASAVLALVFLALGX	FAVQYGN	GEAVQ	MAGGKYIGQLINMYAVTIGGWSRPLV		
		230	240	250	260	270	280
20	orf53a	AFIAFACMYGTTITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFD					
		290	300	310	320	330	340

The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

	1	ATGTCCGAAC	AACATATTTTC	GACTTGGA	AGTAAATCA	ACGCATTGGG
	51	ACCGGGGATT	ATGATGGCTT	CGCGGGCGGT	CGCGGGTTCG	CACCTGATTG
25	101	CCTCGACGCA	GGCGGGCGCG	CTTACGGCT	GGCAGATCGC	GCTCATCATC
	151	ATCCTGACCA	ACCTCTTCAA	ATACCCGTTT	TTCCGCTTCA	CGCGCGATTA
	201	CACGCTGGAC	ACGGGCAAGA	GCCTGATTGA	AGGTTATGCC	GAGAAAAGCC
	251	GCGTTTATTT	GTGGGTATTC	CTGATTTTGT	GCATCCTCTC	CGCCACGATT
	301	AACGCGGGCG	CGGTCGCCAT	TGTAACCGCC	GCCATCGTCA	AAATGGCGAT
	351	TCCTCTCGCTG	ATGTTTGTATG	CCGGCACGGT	TGCCGCTTG	ATTATGGCAT
30	401	CCTGCCTGAT	TATTTTGGTG	AGCGGACGTT	ACCGCGCTTT	GGATCGCGTT
	451	TCCAAAATCA	TCATCGTTAC	TTTGAGTATC	GCCACGCTTG	CCGCCGCCGG
	501	CATCGCTATG	TCGCGCGGTA	TGCAGATGCA	GTCCGATTTT	ATCGAGCCGA
	551	CACCGTGGAC	GCTTGCCGGT	TTGGGCTTCC	TGATCGCGCT	GATGGGCTGG
35	601	ATGCCCCGCG	CGATTGAAAT	TTCCGCCATC	AATCTTTTGT	GGTAACCGA
	651	AAAACAACGC	ATCAATCCTT	CCGAATACCG	CGACGGGAT	TTTGATTTC
	701	ACGTCGGTTA	TATCGCCAGT	GCGGTTTGG	CTTTGGTTT	CCTTGCATG
	751	GGCGCGTTTG	TGCAATACGG	CAACGGCGAA	GCAGTGCAGA	TGGCGGGCGG
	801	CAAATATATC	GGGCAATTGA	TCAATATGTA	CGCCGTTACC	ATCGGCGGCT
40	851	GGTCGCGCCC	GCTGGTGGCG	TTTATCGCGT	TTGCCTGTAT	GTACGGCACG
	901	ACGATTACCG	TTGTGGACGG	CTATGCCCGT	GCCATTGCCG	AACCCGTGCG
	951	CCTGCTGCGC	GGAAAAGACA	AAACGGGCAA	CGCCGAATTC	TTTGCCTGGA
	1001	ATATTTGGGT	GGCGGGCAGC	GGTTTGGCGG	TGATTTTCTG	GTTTGACGCG
	1051	GTAATGGCGA	ATCTGCTCAA	ATTTGCGATG	ATTGCCGCTT	TTGTGTCCGC
45	1101	CCCTGTGTTT	GCCTGGCTGA	ATTACCGTTT	GGTCAAAGGT	GATGAAAAC
	1151	ACAAACTCAC	ATCAGGTATG	AATGCCCTTG	CATTGGCAGG	CTTGATTTAT
	1201	CTGACCGGTT	TTACCGTTT	GTCTTTATG	AATTTGGCGG	GAATGTTCAA
	1251	ATGA				

This encodes a protein having amino acid sequence <SEQ ID 482>:

	1	MSEQHISTWK	SKINALGPGI	MMASAAVGGG	HLIASTQAGA	LYGWQIALII
50	51	ILTNLFKYPF	FRFSAHYTLD	TGKSLIEGYA	EKSRVYLWVF	LILCILSATI
	101	NAGAVAIITA	AIVKMAIPSL	MFDAGTVAAL	IMASCLIIILV	SGRYRALDRV
	151	SKIIIVTL	SIATLAAAGIAM	SRGMQMOSDF	IEPTPWTL	LAGLGLIALMGW
	201	MPAPIEISAI	NSLWVTEKQR	INPSEYRDGI	FDNFVGYIAS	AVLALVFLAL
	251	GAFAVQYGN	GEAVQ	MAGGKYI	QQLINMYAVT	IGGWSRPLVA
55	301	TITVVDGYAR	AIAEPVRLLR	GKDKTGNAEF	FAWNIWVAGS	GLAVIFWFDG
	351	VMANLLKFAM	IAAFVSAFV	AWLNYRLVKG	DEKHKLTSGM	NALALAGLIY
	401	LTGFTVLFL	LLAGMFK*			

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

		10	20	30	40	50	60
60	orf53a.pep	MSEQHISTWKS	KINALGPGIM	MASAAVGGSH	LIAS	TQAGALY	GWQIALII
	orf53-1	MSEQHISTWKS	KINALGPGIM	MASAAVGGSH	LIAS	TQAGALY	GWQIALII
		10	20	30	40	50	60
65		70	80	90	100	110	120

-284-

5	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAAIVKMAIPSL
	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAAIVKMAIPSL
10	orf53a.pep	130 140 150 160 170 180 MFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIAMSRGMQMOSDF
	orf53-1	130 140 150 160 170 180 MFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIAMSRGMQMOSDF
15	orf53a.pep	190 200 210 220 230 240 IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS
	orf53-1	190 200 210 220 230 240 IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS
20	orf53a.pep	250 260 270 280 290 300 AVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
	orf53-1	250 260 270 280 290 300 AVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
25	orf53a.pep	310 320 330 340 350 360 TITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM
	orf53-1	310 320 330 340 350 360 TITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM
30	orf53a.pep	370 380 390 400 410 IAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLNLAGMFKX
	orf53-1	370 380 390 400 410 IAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLNLAGMFKX

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N.*

#### *gonorrhoeae*:

40	orf53.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIA	91
45	orf53.pep	MSRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
	orf53ng	MSRGMQMOPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
50	orf53.pep	IFEFNVGYYIASAVLALVFLALGXVAPNGNGXTVMAGGKYNGQLINMYA	139
	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYNGEAVQMGGGKYIGQLINMYAVTIGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

55	1	MPKKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMOP
	101	DFIEPTPWT LAGLGFLIALM GWMPAPIEIS AINSLWVTEK QRINPSEYRD
	151	GIFDFNVGYI ASAVLALVFL ALGAFVQYGN GEAVQMGGGK YIGQLINMYA
	201	VTIGGSRPL VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRKDKTARP
	251	IVLLEKLGR HRFGRDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1	..aagaAAAGCT GCGTTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
	51	CGCCACGATT AACGCGGGCG CGGTCGCCAT TGTAACCGCC GCCATCGTCA
	101	AAATGGCGAT TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG

-285-

5  
10  
15

```

151  ATTATGGCAT CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT
201  GGATCGTGTT TCCAAAATCA TCATTGTTAC TTTGAGCATC GCCACGCTTG
251  CCGCCGCCGG CATCGCTATG TCGCGCGGTA TGCAGATGCA GCCCGATTTT
301  ATCGAGCCGA CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT
351  GATGGGCTGG ATGCCCCGCG CGATCGAAAT TTCCGCCATC AATTCTTTGT
401  GGGTAACCGA AAAACAACGC ATCAATCCTT CTGAATACCG CGACGGGATT
451  TTCGATTTCa ACGTGCGTTA TATCGCagT GCGGTTTGG CTTTGGTTTT
501  CCTTGCACTG GGC GCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA
551  TGGCGGGCGG CAAATATATC GGGCAATTGA TTAATATGTA TGCCGTAACC
601  ATCGGCGGCT GGTCTCGTCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT
651  GTACGGCACG ACGATTACCG TTGTGGACGG TTATGCGCGT GCCATTGCCG
701  AACCCGTGCG CCTGCTGCGC GGCAGGGATA AAACGGCAA CGCCGAGTTG
751  TtgccTGGa ATATTGGGT GCGGGGACG GGTtGGCGG TGATTTCTG
801  GTTTGACggc gcaaTGGCgG AAcTgcTCAa ATTTGCGATG ATtggcgcCT
851  TTGTGTCCCG CCCTGTGTTC GCCTGGCTCA ACTACGCCT CGTCAAAGGG
901  GACAAACGCC ACAGGCTTAC CGCCGGTATG AACGCCCTTG CCATTGTCGG
951  CCTGCTCTAC CTGGCCGGGT TTGCCGTTTT GTTCCTGTTG AACCTTACCG
1001 GACTTTTGGC ATAG

```

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

20  
25

```

1  ..KKSCVYLWVF LILCIASATI NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL
51  IMASCLILV SGRYRALDRV SKIIIVTLSI ATLAAAGIAM SRGMQMOPDF
101 IEPTPWTLag LGFLIALMGW MPAPIEISAI NSLWVTEKQR INPSEYRDI
151 FDFNVGYIAS AVLALVFLAL GAFVQYNGE AVQMAGGKYI GQLINMYAVT
201 IGGWSRPLVA FIAFACMYGT TITVVDGYAR AIAEPVRLLR GRDKTGNAEL
251 FAWNIVVAGS GLAVIFWFDG AMAELLKFAM IAAFVSAPVF AWLNYRLVKG
301 DKRHRLTAGM NALAIVGLLY LAGFAVLFL NLTGLLA*

```

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

30  
35  
40  
45  
50  
55  
60  
65

```

               60      70      80      90      100     110
orf53-1.pep  ILTNLFKYPFFRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTA
orf53ng-1      :|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
               10      20      30
               120     130     140     150     160     170
orf53-1.pep  AIVKMAIPSLMFDAGTVAALIMASCLILVSGRYRALDRVSKIIIVTLSIATLAAAGIAM
orf53ng-1    AIVKMAIPSLMFDAGTVAALIMASCLILVSGRYRALDRVSKIIIVTLSIATLAAAGIAM
               40      50      60      70      80      90
               180     190     200     210     220     230
orf53-1.pep  SRGMQMOSDFIEPTPWTLagLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDI
orf53ng-1    SRGMQMOPDFIEPTPWTLagLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDI
               100     110     120     130     140     150
               240     250     260     270     280     290
orf53-1.pep  FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA
orf53ng-1    FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA
               160     170     180     190     200     210
               300     310     320     330     340     350
orf53-1.pep  FIAFACMYGTTITVVDGYARAIAEPVRLLRGDKDTGNAEFFAWNIVVAGSGLAVIFWFDG
orf53ng-1    FIAFACMYGTTITVVDGYARAIAEPVRLLRGDKDTGNAELFAWNIVVAGSGLAVIFWFDG
               220     230     240     250     260     270
               360     370     380     390     400     410
orf53-1.pep  VMANLLKFAMIAAFVSAPVFAWLNRYLVKGDKEHKLTSGMNALALAGLIYLTGFTVLFLL
orf53ng-1    AMAELLKFAMIAAFVSAPVFAWLNRYLVKGDKRHRLTAGMNALAIVGLLYLAGFAVLFL
               280     290     300     310     320     330
orf53-1.pep  NLAGMFKX
orf53ng-1    ||:|::
orf53ng-1    NLTGLLAX

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1  ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGT
51  TGGCCTTGCC GGCTTGT TTTT TGTCCCGCGC ACAATCCGAA CGCGAGTGG
101 TGGCGGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG
151 CTGCCTGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT
201 GCTTTTCCAC GCCGTCAAAA CGGCAGTGTA TTGGCTGTTT GTCGGTGTGG
251 TCCGTTTCTG CCGAACTAT CTGGCGCACG AATCCGAACC GGACAGGCC
301 GTTCCGCCT..

```

This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

1  ..LRETAYVLDs FDRYFVVALA GLFFVRAQSE REWMREVSaw QEKKGKQAE
51  LPEIKDGMpD FPELALMLFH AVKTAVYWLF VGVVRFRCrNY LAHESEpDRP
101  VPP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

20  1  ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTTGTG CGCGCACAAAT CCGAACCGCA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTGGC TGTTTGTCGG TGTCGTCCGT TTCTGCCGAA
25  251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGGAAGA AGCAGAAGCT GCGGAGGAAG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
30  501 AATTTCGCCC GTCGTCGGG TTTTAAAGA AATCACTTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTAAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCCATATG TTCGATGCGG ACAAGAAGCT GTTTTCCGAG
35  751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTCTGCA GTCAAAGCCG AAAATGCAGG GAATGCGCGG TTCCACCGTC
851 ATGCAGGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACGGCC GTCCGCGATG CCGGCCGCGG
951 CGTTTCCGTC AATTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
40  1001 CGCGAATTTT TCGCTGATG CCGGAAAGTC AGACGGTTGT CCGGAAACGG
1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTACGG AAACCGTTTC
1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAACTGCC GATATCCATA
1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
1201 CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCCGC CTCCCGTATC
45  1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GTCAGGATTC GAGCAGGTGC
1301 AACGCAGCGG CATTGCCGAG ACCGACCATC TTGCCGATGA TGTTTGAAT
1351 GGAGGTGGC AGGAGGAAAC CGCCGCTATT GCGGATGACG GCAGTGAAGG
1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAACCC GAAGCGTTTCG
1451 GGCATGACAG TCAGGCGGTT TGTCCGTTG AAAATGTGCC GTCTGAACGC
50  1501 CCGTCTGCC GGTATCGGA TACGGAAGCG GATGAAGGG CGTTCCCATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG
1651 GAAAAACAGC TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
701 CAAGGTTGTC GATTCTTATT CCGGCCCGCT AATTACGCGT TATGAAATCG
55  1751 AACCCGATGT CGGCGTGCGC GGCAATTCCG TTCTGAATCT GAAAAAGAT
1801 TTGCGCGCTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCT AAACCATCCC
1851 CCGCAAAACC TGCATGGGTT TGGAACTTCC GAACCCGAAA CGCCAAATGA

```

1901 TACGCCTGAG CGAAATCTTC AATTCGCCCG AGTTTGCCGA ATCCAAATCC  
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC  
 2001 CGACTTGGGA AAAGCACCGC ATTTGTTGGT TGCCGGCACG ACCGGTTCGG  
 5 2051 GCAAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC  
 2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT  
 2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCTGTCT GTTACCGATA  
 2201 TGAAGCTGGC GGCAACGCG CTGAACCTGG GTGTTAACGA AATGGAAAAA  
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGTAATCTTG CGGGCTTCAA  
 10 2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT  
 2351 TCAGCCTCAC GCGGACGAT CCCGAACCTT TGGAAAACT GCCGTTTATC  
 2401 GTGGTCGTGG TCATGAGTAT TGCCGACCTG ATGATGACGG CAGGCAAGAA  
 2451 AATCGAAGAA CTGATTGCCG GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA  
 2501 TCCATTGAT TCTTGCCACA CAACGCCCA GCGTCGATGT CATCACGGGT  
 15 2551 CTGATTAAAG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA  
 2601 AATCGACAGC CGCACGATTC TCGACCAAT GGGCGCGGAA AACCTGCTCG  
 2651 CTCAGGGCGA TATGCTGTTC CTGCTGCCGG GTACTGCCTA TCCGCAGCGC  
 2701 GTTCACGGCG CGTTTGCTTC GGATGAAGAG GTGCACCGCG TGGTCGAATA  
 2751 TTTGAAACAG TTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGGCG  
 2801 GCGGCAGCGA AGAGCTGCCG GGCATCGGGC GCAGCGCGCA CGACGAAACC  
 20 2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCTTAAAA CGCGCAAGAC  
 2901 CAGCATTTTC GGCGTACAGC GCGCCTTGCG TATCGGCTAC AACC CGCGCG  
 2951 CCGCTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA  
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

25 1 MFWIVLIVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK  
 51 DGMPPDFPELA LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRPVPPAS  
 101 ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEAADTED IATAVIDNRR  
 151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI  
 201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE  
 30 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQCKG QAEAKSPDVS  
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR  
 351 DVEMPSETEN VFTETVSSVG YGCPVYDETA DIHIEEPAAP DAWVVEPPEV  
 401 PKVPMTAIDI QPPPPVSEIY NRTYEPSPGF EQVQSRIRAE TDHLADVDIN  
 451 GGWQEETAAI ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CPFENVPSER  
 35 501 PSCRVSDETA DEGAFFSEET GAVSEHLPTT DLLLPLFNP EATQTEELL  
 551 ENSITIEEKL AEFKVKVKVV DSYSGPVITR YEIEPDVGVR GNSVLNLEKD  
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS  
 651 KLTLALGQDI TGQPVVTDLG KAPHLLVAGT TSGSKSVGVN AMILSMLFKA  
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMLAANA LNWCVNEMEK  
 40 751 RYRLMSFMGV RNLAFNQKI AEAAARGEKI GNPFSLTPDD PEPLEKLPFI  
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG  
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LLPGTAYPQR  
 901 VHGAFAFASDEE VHRVVEYLKQ FGEPDYVDDI LSGGGSEELP GIGRSGDDET  
 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAFE  
 45 1001 HNGNRTILVP LDNA\*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N.*

50 *meningitidis*:

10 20 30 40 50 60  
 orf58.pep LRETAYVLDSFDTRYFVVALAGLFFVRAQSEREWMRVSAWQEKKGKQAELEPEIKDGMPPD  
 orf58a MFWIVLIVILLLALAGLFFVRAQSEREWMRVSAWQEKKGKQAELEPEIKDGMPPD  
 10 20 30 40 50  
 70 80 90 100  
 orf58.pep FPELALMLFHAVKTA VYWLFGVVVRFCRNYLAHES EPDRPVPP  
 orf58a FPELALMLFHAVKTA VYWLFGVVVRFCRNYLAHES EPDRPVPPASANRADVPTASDGYS  
 60 70 80 90 100 110

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

1  ATGTTTGGG TAGTTTGGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTGTC CGCGCACAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101  CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAATCAAA
5  151  GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201  CAAAACGGCA GTGTATTGGC TGTGTGTCGG TGTGTCCTGT TTCTGCCGAA
251  ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301  GCAAATCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351  AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
10  401  AGGTGCGCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451  ATCCATTTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
501  AATTTCGCCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
551  CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601  GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
15  651  TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701  AACGCACGTA TTCCCGTATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG
751  TTGCGGATT ACAGATTGTA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801  CTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCCGTC
851  ATGCAGGGCA GGGNAAAGGG CAGGCGGAGG CNAATCCCC GGATGTTTCC
20  901  CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCNGCCGCCG
951  CGTTTCCGCT AATTGAAAG AACCGAACAA GGCACCGGTT TCTCGCGAGG
1001  CGCGGATTTT GCGCTGATT CCGGAAAGTC GGACGTTTGT CCGGAAACGG
1051  GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTCACGG AAANTGTTTC
1101  GTCTGTGGGA TACGGCGNTC CGGTTTATGA TGAAGTCCG GATATCCATA
25  1151  TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCAACC ACCCGAAGTG
1201  CCGAAAGTTC CCATGCCCCG AATNGATATT CCGCCGCCGC CTCCCGTATC
1251  GGAATCTAC AACCGTACCT ATGAACCGCC GGCAGGATTC GAGCAGGTGC
1301  AACGCAGCCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTGAAT
1351  GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG GCAGTGAGG
30  1401  TGTGGCAGAG CGGTCAAGCG GGCAATATTT GTCGGAAACC GAAGCGTTCG
1451  GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
1501  CCGTCCCGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
1551  TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601  TGCCGCGGCT GTTCAATCCC GGGCGGACGC AAACCGAAGA AGANCTGTTG
35  1651  GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1701  CAAGGTTGTC GATTCTTAT CCGGCCCGGT GATTACGCGT TATGAAATCG
1751  AACCCGATGT CCGCGTGCGC GGCAATTCCG TTCTAAATCT GGAAGAAAGAN
1801  TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCG AAACCATCCT
1851  CGGCAAAACC TGTATGGGTT TGGAACTTCC GAACCCGAAA CGCCAAATGA
40  1901  TACGCTGAG CGAAATCTTC AATTGCGCCG AGTTTGCCGA ATCCAAATCC
1951  AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAC
2001  CGACTTGGGC AAAGCACC GC ATTTGTTGGT TGCCGGCAGC ACCGGTTCGG
2051  GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
45  2101  GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151  GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCTGTC GTTACCGATA
2201  TGAAGCTGGC GGCAAACGCG CTGAACTGGT GTGTTAACGA AATGGAAGAAA
2251  CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CCGGTNTCAA
2301  TCAAAAAATC GCCGAAGCCG CAGCAAGGGG GGAGAAAAATC GGCAACCCGT
50  2351  TCAGCCTCAC GCCGACAAT CCCGAACCTT TGGANAAATT GCCGTTTATC
2401  GTGTCGTGG TTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
2451  AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
55  2501  TCCATCTTAT CCTTGCCACA CAACGCCCCA GTGTCGATGT CATCACGGGT
2551  CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
2601  AATCGACAGC CGCAGGATTC TTGACCAAAT GGGTGCAGGA AACCTGCTCG
2651  GGCAGGGCGA TATGCTGTTT CTGCCGCCGG GTACGGCCTA TCCGCAGCGC
2701  GTTCACGGCG CGTTGCTTC GATGAAGAG GTGCACCGCG TGGTCAATA
2751  TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTGAGCGGCG
2801  GTATGTCCGA CGATTGCTG GGAATCAGCC GGAGCGGCGA CGGCGAAACC
60  2851  GATCCGATGT ACGACGAGGC CGTGTGNGTT GTTTTGAAAA CGCGCAAAGC
2901  CAGCATTTCT GCGGTGCGAG CCGCATTCG TATCGGCTAT AATCGCGCGC
2951  CGGTCTGAT TGACGAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
3001  CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

```

1  MFWIVLVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
65  51  DGMPDPPELA LMLFHAVKTA VYWLFGVVR FCRNYLAHES EPDRPVPPAS
101  ANRADVPAS DGYSDSNGT EEAETEEAEA AEEEAADTED IATAVIDNRR
151  IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
201  DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSRM FDADKEAFSE

```

	251	SADYGFEPEYF	EKQHPSAFSA	VKAENARNAP	FRRHAGQGKG	QAEAKSPDVS
	301	QGQSVSDGTA	VRDAXRRVSV	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
	351	DVEMPSETEN	VFTXVSSVG	YGXPVYDETA	DIHIEPAAP	WDWVVEPEV
5	401	PKVMPAXDI	PPPPVSEIY	NRTYEPAGF	EQVQRSRIAE	TDHLADDVLN
	451	GGWQEETAAI	ANDGSEGVAE	RSSGQYLSET	EAFGHDSQAV	CPFENVPSER
	501	PSRRAXDTEA	DEGAFQSEET	GAVSEHLPTT	DLLPPLFNP	GATQTEEXLL
	551	XNSITIEEKX	AEFKVKVKV	DSYSGPVITR	YEIEPDVGV	GNSVLNLEKX
	601	LARSLGVASI	RVVETILGKT	CMGLELNP	RQMIRLSEIF	NSPEFAESKS
10	651	KLTLALGQDI	TGQPVVTDLG	KAPHLVAGT	TGSGKSVGVN	AMILSMLFKA
	701	APEDVRMIMI	DPKMLELSIY	EGIPHLLAPV	VTDMKLAANA	LNWCVNEMEK
	751	RYRLMSFMGV	RNLGXNQKI	AEAAARGEKI	GNPFSLTDPN	PEPLXKLPTI
	801	VVVVDEFADL	MMTAGKKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
	851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLQGQDMLF	LPPGTAYPQR
	901	VHGAFASDEE	VHRVVEYLKQ	FGEPTYVDDX	LSGMSDDL	GISRSGDGET
15	951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEGIVSAPE
	1001	HNGNRTILVP	XDNA*			

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

		10	20	30	40	50	60
20	orf58a.pep	MFWIVLIVILLALAGLFFVRAQSEREWREVSAWQEKKGKQAE	LPKIDGMPDFPELA				
	orf58-1	MFWIVLIVILLALAGLFFVRAQSEREWREVSAWQEKKGKQAE	LPKIDGMPDFPELA				
		10	20	30	40	50	60
25	orf58a.pep	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPD	RPVPPASANRADVPTASDGYSDSGNGT				
	orf58-1	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPD	RPVPPASANRADVPTASDGYSDSGNGT				
		70	80	90	100	110	120
30	orf58a.pep	EEAEETEEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEIS	SPVRPVFKEITL				
	orf58-1	EEAEETEEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEIS	SPVRPVFKEITL				
		130	140	150	160	170	180
35	orf58a.pep	EEATRALNSAALRETKKRYIDAFEKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSRM					
	orf58-1	EEATRALNSAALRETKKRYIDAFEKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSHM					
		190	200	210	220	230	240
40	orf58a.pep	FDADKEAFSESADYGFEPEYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS					
	orf58-1	FDADKEAFSESADYGFEPEYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS					
		250	260	270	280	290	300
45	orf58a.pep	QGQSVSDGTAVRDAXRRVSVNLKEPNKATVSAEARISRLIPESRTVVGKR	DVEMPSETEN				
	orf58-1	QGQSVSDGTAVRDAXRRVSVNLKEPNKATVSAEARISRLIPESRTVVGKR	DVEMPSETEN				
		310	320	330	340	350	360
50	orf58a.pep	VFTXVSSVGYGXPVYDETA	DIHIEPAAPDAWVVEPEV	PKVMPAXDIPPPPVSEIY			
	orf58-1	VFTXVSSVGYGXPVYDETA	DIHIEPAAPDAWVVEPEV	PKVMPAXDIPPPPVSEIY			
		370	380	390	400	410	420
55	orf58a.pep	NRTYEPAGFEQVQRSRIAE	TDHLADDVLNGGWQEETA	AIANDGSEGVAE	RSSGQYLSET		
	orf58-1	NRTYEPAGFEQVQRSRIAE	TDHLADDVLNGGWQEETA	AIANDGSEGVAE	RSSGQYLSET		
		430	440	450	460	470	480
60	orf58a.pep	EAFGHDSQAVCPFENVPSERPSRRAXDTEA	DEGAFQSEETGAVSEHLPTD	LLPPLFNP			
	orf58-1	EAFGHDSQAVCPFENVPSERPSRRAXDTEA	DEGAFQSEETGAVSEHLPTD	LLPPLFNP			
		490	500	510	520	530	540
65	orf58a.pep	EAFGHDSQAVCPFENVPSERPSRRAXDTEA	DEGAFQSEETGAVSEHLPTD	LLPPLFNP			
	orf58-1	EAFGHDSQAVCPFENVPSERPSRRAXDTEA	DEGAFQSEETGAVSEHLPTD	LLPPLFNP			
		490	500	510	520	530	540
70	orf58a.pep	EAFGHDSQAVCPFENVPSERPSRRAXDTEA	DEGAFQSEETGAVSEHLPTD	LLPPLFNP			
	orf58-1	EAFGHDSQAVCPFENVPSERPSRRAXDTEA	DEGAFQSEETGAVSEHLPTD	LLPPLFNP			

-290-

		550	560	570	580	590	600
	orf58a.pep	GATQTEEXLLXNSITIEEKXAEFKVKVVDVSYSGPVITRYEIEPDVGVRGNSVLNLEKX					
5	orf58-1	EATQTEEELENSITIEEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVRGNSVLNLEKD					
		550	560	570	580	590	600
	orf58a.pep	LARSLGVASIRVVETILGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI					
10	orf58-1	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI					
		610	620	630	640	650	660
	orf58a.pep	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
15	orf58-1	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
		670	680	690	700	710	720
	orf58a.pep	EGIPHLLAPVVTDMLAANALNWCNEMEKRYRLMSFMGVRNLAGXNQKIAEAAARGEKI					
20	orf58-1	EGIPHLLAPVVTDMLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58a.pep	GNPFSLTPDNPEPLXKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
25	orf58-1	GNPFSLTPDDPEFLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58a.pep	QRPSVDVITGLIKANIPTRIAFOVSSKIDSRITLDQMGAEENLLGQGDMLFLPGTAYPQR					
30	orf58-1	QRPSVDVITGLIKANIPTRIAFOVSSKIDSRITLDQMGAEENLLGQGDMLFLPGTAYPQR					
		850	860	870	880	890	900
	orf58a.pep	VHGAFASDEEVHRVVEYLKQFGEPDYDDXLGGMSDDLGISRSGDGETDPMYDEAVSV					
35	orf58-1	VHGAFASDEEVHRVVEYLKQFGEPDYDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58a.pep	VLKTRKASISGVQRALRIGYNRAARLIDOMEAEGIVSAPEHNGNRTILVLPXDNAX					
40	orf58-1	VLKTRKASISGVQRALRIGYNRAARLIDOMEAEGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	
	orf58a.pep	VLKTRKASISGVQRALRIGYNRAARLIDOMEAEGIVSAPEHNGNRTILVLPXDNAX					
45	orf58-1	VLKTRKASISGVQRALRIGYNRAARLIDOMEAEGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

	orf58.pep	ALMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDRPVPP	103
55	orf58ng	SEPDRPVPPASANRADVPTASDGYSDSGNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

	1	..SEPDRPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAETEAAE	AAEEEEADTE
60	51	DIATAVIDNR	RIPFDRSIAE	GLMQSESKTS	PVRPVFKEIT	LEEATRALSS
	101	AALRETKKRY	IDAFEKNGTA	VPKVRVSDTP	MEGLQIIIGLD	DPVLQRTYSR
	151	MFDADKEAFS	ESADYGFEPY	FEKQHPSAFS	AVKAENARNA	PFRRHAGQEK
	201	GQAEAKSPDV	SQGQSVSDGT	AVRDARRRVS	VNLKEPNKAT	VSAEARISRL
	251	IPESRTVVGK	RDVEMPSETE	NVFTETVSSV	GYGGPVYDEA	ADIHIEEPAA
	301	PDAAVVVEPPE	VPEVAVPEID	ILPPPPVSEI	YNRTYEPPAG	FEQAQRSRIA



351	ETDHLAADVL NGGWQEETAA IADDGSEGAA ERSSGQYLSE TEAFGHDSQA
401	VCPFEDVPSE RPSRVSDTE ADEGAFAQSEE TGAVSEHLPT TDLLLPPLFN
451	PEATQTEEL LENSITIEEK LAEFKVKVKV VDSYSGPVIT RYEIEPDVGV
501	RGNSVLNLEK DLARSLGVAS IRVETIPGK TCMGLELPNP KRQMIRLSEI
551	FNSPEFAESK SKLTLALGQD ITGQPVVTDL GKAPHLVAG <u>TTGSGKSVGV</u>
601	NAMILSMLFK AAPEDVRMIM IDPKMLELSI YEGITHLLAP VVTDMKLAAN
651	ALNWCVNEME KRYRLMSFMG VRNLAGFNQK IAEAAARGEK IGNPFSLTPD
701	DPEPLEKLPF <u>IVVVVDEFAD</u> LMMTAGKKIE ELIARLAQKA RAAGIHLILA
751	TQRPESVDVIT GLIKANIPTR IAFQVSSKID SRITLDQMGAE ENLLGQGDML
801	FLPPGTAYPQ RVHGAFASDE EVHRVVEYLK QFGEPTYVDD ILSGGGSEEL
851	PGIGRSGDGE TDPMYDEAVS VVLKTRKASI SGVQRALRIG YNRAARLIDQ
901	MEAEIGIVSAP EHNGNRTILV PLDNA*

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng:	467	IEEKLAEFKVKVKVVDYSYSGPVITRYEIEPDVGV	RGNSVLNLEKDLARSLGVASIRVVET	526
		+E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE		
FtsK:	868	VEARLADFRIKADVNYSPGPVITRFELNLAGPVKAARISNLSRDLARSLSTVAVRVVEV		927
ORF58ng:	527	IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTLALGQDITGQPVVTDLGKAPHL		586
		IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL		
FtsK:	928	IPGKPYVGLELPNKKRQTVYLRVLDNAKFRDNPSPLTVVLGKDIAGEVVDLAKMPHL		987
ORF58ng:	587	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK		646
		LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK		
FtsK:	988	LVAGTTGSGKSVGVNAMILSMYKAQPEDVRFIMIDPKMLELSVYEGIPHLLEVVVTDMK		1047
ORF58ng:	647	LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP--		704
		AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +		
FtsK:	1048	DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPIDPDYKPGDSMDAQH		1107
ORF58ng:	705	--LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPESVDVITGL		762
		L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPESVDVITGL		
FtsK:	1108	PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLTATQRPESVDVITGL		1167
ORF58ng:	763	IKANIPTRIAFQVSSKIDSRITLDQMGAEENLLGQGDMLFLPPGTAYPQRVHGAFASDEEV		822
		IKANIPTRIAF VSSKIDSRITLDQ GAE+LLG GDML+ P + P RVHGAF D+EV		
FtsK:	1168	IKANIPTRIAFTVSSKIDSRITLDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV		1227
ORF58ng:	823	HRVVEYLKQFGEPTYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG		882
		H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG		
FtsK:	1228	HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQVTEKRKASISG		1286
ORF58ng:	883	VQRALRIGYNRAARLIDQMEAEIGIVSAPEHNGNRTILVP		921
		VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P		
FtsK:	1287	VQRQFRIGYNRAARIIEQMEAQGVISEQGHNGNREVLAP		1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1	ATGTTTTGGA TAGTTTTGAT CGTTATgtg TTGCTTGCGC TTGCCGCGCT
51	GTTTTTTGTC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101	CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAATCAAA
151	GACGGTATGC CCGATTTTCC CGAGTTTTC CTGATGCTTT TCCATGCCGT
201	CAAAACGGCA GTGTATTGGC TGTTTGTCCG TGTCGTCCGT TTCTGCCGAA
251	ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301	GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGGTATT CAGACAGTGG
351	AAACGGGACG GAAGAAGCGG AAACGGAAGC AGCAGAAGCT GCGGAGGAAG
401	AGGCTGCCgA TACgGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451	ATCCcatTCG ACCGGAGTAT TGCTGAAGGG TTGATGCAGT CTGAAAGCAA
501	AACTTCGCCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
551	CGCGTGCTTT AAGCAGCGCG GCTTAAAGGG AAACGAAAAA ACGCTATATC
601	GATGCATTTG AGAAAAACGG AACAGCCGTC CCCAAAGTAC GCGTGTCCGA
651	TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701	AACGCACGTA TTCCCGTATG TTTGATGCGG ACAAAGAAGC GTTTTCCGAG
751	TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC

	801	CTTTTCTGCA	GTCAAAGCCG	AAAATGCACG	GAATGCGCCG	TTCCGCCGTC
	851	ATGCAGGGCA	GGAGAAAGGG	CAGGCGGAGG	CAAAATCCCC	GGATGTTTCC
	901	CAAGGGCAGT	CCGTTTCAGA	CGGCACAGCC	GTCCGCGATG	CCCGCCGCCG
5	951	CGTTTCCGTC	AATTTGAAA	AACCGAACAA	GGCAACGGTT	TCTGCGGAGG
	1001	CGCGGATTTC	GCGCCTGATT	CCGGAAGTTC	GGACGGTTGT	CGGGAACGCG
	1051	GATGTCGAAA	TGCCGTCTGA	AACCGAAAAT	GTTTTACGG	AAACCGTTTC
	1101	GTCTGTGGGA	TACGGCGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
	1151	TTGAAGAGCC	TGCCGCGCCC	GATGCTTGGG	TGGTCGAACC	ACCCGAAGTG
10	1201	CCGGAGGTAG	CCGTACCCGA	AATCGATATT	CTGCCGCCGC	CTCCCGTATC
	1251	GGAAATCTAC	AACCGTACCT	ATGAGCCGCC	GGCAGGATTC	GAGCAGGCGC
	1301	AACGCAGCCG	CATTGCCGAA	ACCGACCATC	TTGCCGCTGA	TGTTTTGAAT
	1351	GGAGGTGGC	AGGAGGAAAC	CGCCGCTATT	GCAGATGACG	GCAGTGAGGG
	1401	TGCGGCAGAG	CGGTCAAGCG	GGCAATATCT	GTCCGAAACC	GAAGCGTTGC
	1451	GGCATGACAG	TCAGGCGGTT	TGTCCGTTTG	AAGATGTGCC	GTCTGAACGC
15	1501	CCGTCTGCC	GGTATCGGA	TACGGAAGCG	GATGAAGGGG	CGTTCCAATC
	1551	GGAAGAGACC	GGTCCGGTAT	CCGAACACCT	GCCGACAACC	GACCTGCTTC
	1601	TGCTCCGTC	GTTCAATCCC	GAGGCGACGC	AAACCGAAGA	AGAAGCTGTTG
	1651	GAAAACAGCA	TCACCATCGA	AGAAAAATTG	GCGGAGTTCA	AAGTCAAGGT
20	1701	CAAGGTTGTC	GATTCCTTAT	CCGGCCCCGT	GATTACGCGT	TATGAAATCG
	1751	AACCCGATGT	CGGCGTGCGC	GGCAATTCCG	TTCTGAATTT	GGAAAAAGAC
	1801	TTGGCGCGTT	CGCTCGGCGT	GGCTTCCATC	CGCGTTGTGC	AAACCATCCC
	1851	CGGCAAAACC	TGCATGGGTT	TGGAACCTCC	GAACCCGAAA	CGCCAAATGA
	1901	TACGCCTGAG	CGAAATTTTC	AATTCGCCCG	AGTTTGCCGA	ATCCAAATCC
25	1951	AAGCTGACGC	TCGCGTCCGG	TCAGGACATT	ACCGGACAGC	CCGTCTGTAAC
	2001	CGACTTGGGC	AAAGCACCGC	ATTTGCTGGT	TGCCGGCACG	ACCGGTTCCGG
	2051	GCAATTCGGT	GGGTGTCAAC	GCGATGATTC	TGTCTATGCT	TTTCAAAGCC
	2101	GCGCCGGAAG	ACGTGCGTAT	GATTATGATC	GATCCGAAAA	TGCTGGAATT
	2151	GAGCATTTAC	GAAGGCATCA	CGCACCTGCT	CGCCCTGTC	GTTACCGATA
30	2201	TGAAGCTGGC	GGCAAACGCG	CTGAACTGGT	GTGTTAACGA	AATGGAAAAA
	2251	CGTACCGCC	TGATGAGCTT	TATGGGCGTG	CGCAATCTTG	CGGGCTTCAA
	2301	CCAAAAAATC	GCCGAAGCCG	CAGCAAGGGG	AGAAAAAATC	GGCAATCCGT
	2351	TCAGCCTCAC	GCCCGACGAT	CCGGAACCTT	TGGAAAAACT	GCCGTTTATC
	2401	GTGGTCGTGG	TCGATGAGTT	TGCCGATTTG	ATGATGACGG	CAGGCAAGAA
35	2451	AATCGAAGAA	CTGATTGCGC	GCCTCGCCCA	AAAAGCCCGC	GCGGCAGGCA
	2501	TCCACCTTAT	CCTTGCCACA	CAACGCCCCA	GCGTCGATGT	CATCACGGGT
	2551	CTGATTAAGG	CGAAGCATCC	GACGCGTATC	GCGTTCCAAG	TGTCCAGCAA
	2601	AATCGACAGC	CGCACGATTC	TCGACCAAAT	GGGCGCGGAA	AACCTGCTCG
	2651	GTGAGGGCGA	TATGCTGTTC	CTGCCGCCGG	GTAAGCCTA	TCCGCAGCGC
40	2701	GTTACGCGCG	CGTTTGCCCTC	GGATGAAGAG	GTGCACCGCG	TGGTCGAATA
	2751	TCGTAAGCAG	TTTGGCGAGC	CGGACTATGT	TGACGATATT	TTGAGCGCGG
	2801	GCGGCAGCGA	AGAGCTGCCC	GGCATCGGGC	GCAGCGGCGA	CGGCGAAACC
	2851	GATCCGATGT	ACGACGAGGC	CGTATCCGTT	GTCTGAAAA	CGCGCAAAGC
	2901	CAGCATTTTC	GGCGTACAGC	GCGCCTTGCG	CATCGGCTAC	AACCGCGCCG
45	2951	CGGCTCTGAT	TGACCAAATG	GAAGCGGAAG	GCATTGTGTC	CGCACCGGAA
	3001	CACAACGGCA	ACCGTACGAT	TCTCGTCCCC	TTGGACAATG	CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

	1	MFWIVLIVIV	LLALAGLFFV	RAQSEREWMR	EVS AWQEKKG	EKQ AELPEIK
	51	DGMPDFPEFS	LMLEHAVKTA	VYWLFGVVR	FCRNYIAHES	EPDRPVPPAS
50	101	ANRADVPTAS	DGYSDSGNGT	EEAEETAAEA	AEEEEADTED	IATAVIDNRR
	151	IPFDRSIAEG	LMQSESKTSP	VRPVFKEITL	EEATRALSSA	ALRETKKRYI
	201	DAFEKNGTAV	PKVRVSDTPM	EGIQIIGLDD	PVLQRTYSRM	FDADKEAFSE
	251	SADYGFEPYF	EKQHPSAFSA	VKAENARNAP	FRRHAGQKEG	QAEAKSPDVS
	301	QQQSVSDGTA	VRDARRRVSV	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
55	351	DVEMPSETEN	VFTETVSSVG	YGGFVYDEAA	DIHIEEPAP	DAWVVEPEV
	401	PEVAVPEIDI	LPPFPVSEIY	NRTYEPPAGF	EQAQRSRIAE	TDHLAADVLN
	451	GGWQEETAAI	ADDGSEGAAE	RSSGQYLSET	EAFGHDSQAV	CPFEDVPSE
	501	PSCRVSDETA	DEGAFQSEET	GAVSEHLPTT	DLLEPLFPNP	EATQTEEELL
	551	ENSITIEEKL	AEPKVVKVSV	DSYSGPVITR	YEIEPDVGV	GNSVLNLEKD
60	601	LARSLGVASI	RVVETIPGKT	CMGLELPNPK	RQMIRLSEIF	NSPEFAESKS
	651	KLTLALGQDI	TGQPVVTDLG	KAPHLLVAGT	TSGSKSVGVN	AMILSMLFKA
	701	APEDVRMIMI	DPKMLELSIY	EGITHLLAPV	VTDMKLAANA	LNWCVNEMEK
	751	RYRLMSFMGV	RNLAGFNQKI	AEEAARGEKI	GNPFSLTDD	PEPLEKLPFI
	801	VVVVDEFADL	MMTAGKKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
65	851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQGDMLF	LPPGTAYPQR
	901	VHGAFASDEE	VHRVVEYLKQ	FGEPTYDDI	LSGGGSEELP	GIGRSGDGET
	951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEIGVSAPE
	1001	HNGNRTILVP	LDNA*			

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

10

15

20

25

30

35

40

45

50

55

65

70

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
5		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
10		790	800	810	820	830	840
	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR					
15		850	860	870	880	890	900
	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
20		910	920	930	940	950	960
	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPLDNAX					
25		970	980	990	1000	1010	
	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

30 Furthermore, ORF58ng-1 shows significant homology to the *E.coli* protein FtsK:

	sp P46889 FTSK_ECOLI CELL DIVISION PROTEIN FTSK >gi 1651412 gnl PID d1015290 (D1 division protein FtsK [Escherichia coli] >gi 1651418 gnl PID d1015296 (D90727) Cell division protein FtsK [Escherichia coli] >gi 1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329
35	Score = 576 bits (1469), Expect = e-163 Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)
	Query: 556 IEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 615 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE
40	Sbjct: 868 VEARLADFRIKADVNNYSPGPVITRFEFLNLAGPVKAARISNLSRDLARSLSTVAVRVVEV 927
	Query: 616 IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL 675 IPGK +GLELEN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL
45	Sbjct: 928 IPGKPYVGLELPNKKRQTVYLVREVLNDAKFRDNPSPLTVVLGKDIAAGEPVVADLAKMPHL 987
	Query: 676 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 735 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK
	Sbjct: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVTDMK 1047
50	Query: 736 LAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP-- 793 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +
	Sbjct: 1048 DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPIDPDYWKPGDSMDAQH 1107
	Query: 794 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPVSDVITGL 851 L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPVSDVITGL
55	Sbjct: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPVSDVITGL 1167
	Query: 852 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQRVHGAFASDEEV 911 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV
60	Sbjct: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227
	Query: 912 HRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 971 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG
65	Sbjct: 1228 HAVVQDWKARGRPQYVDGITSDSESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286
	Query: 972 VQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVP 1010 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P
	Sbjct: 1287 VQRQFRIGYNRAARIIEQMEAQIVSEQHNGNREVLP 1325

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

5      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CCGTATTGGT CTCCACGCAG GCAATCAACC
     101  TGCTCGGCCG TGCCGCCGAC GGGC...GTGA TCGCCATCGA TGCCGTGTTG
     151  GCATTGGTCG GCTTCTGGGT C.....
          //
    10      901  .....A TTGCCATCGG TTTGTTTTTA ATTTACCAAA ACGGGCTGAC
          951  CCTGCTTTT GAAGCCGTGG AAGACGGCAA AATCCATTTT TGGCTCGGAC
     1001  TGCTGCCTAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTTGCGC
     1051  GTCCGCAGTA TGCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAAAGTCT
     1101  GACATTGAAA GGCGGAAAT GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL
     51  ALVGFVW... .....
          //
    301  ...IAIGLFL IYQNGLTLLF EAVEDGKIHF WLGLLPMHII MFVLALILLR
    351  VRSMPSQPFW QAVGKSLTLK GK*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
     51  CATTTTCGTC GTCCTCTTGG CCGTATTGGT CTCCACGCAG GCAATCAACC
    101  TGCTCGGCCG TGCCGCCGAC GGGCGTGTG CCGTATTGGT CCGTATTGGT
    151  TTGGTCGGCT TCTGGGTCAT CCGTATGACG CCGCTTTTGC TGGTGTGAC
    201  CGCATTATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CCGACAGCG
    251  AAATGTCGGT CTGGCTATCC TGCGGATTGG CATTGAAACA ATGGATACGC
    301  CCGGTGATGC AGTTTGCCGT GCGGTTGCC GTTTTGGTTG CCGTCATGCA
    351  GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
    401  TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTTCAAC
    451  AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTGAAA CCTTCGATAC
    501  CGAATCCGGC ATCATGAAAA ACCTGTTTCT GCGCGAACAG GACAAAAACG
    551  CGGGCGACAA CATCATCTTC GCCAAGAAG GTAACCTCTC GCTGAACGAC
    601  AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
    651  CCGACGCGCC GACTACAATC AGGTTTCTCT CCAAAACTC AACCTGATTA
    701  TCAGCACCAC GCCCAACTC ATCGACCCCG TTTCCACCG CCGTACCATT
    751  CCGACCGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
    801  GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCCTGCTTG
    851  CCGTGCCGCT TTCTATTTT AACCCTGCGC GCGGACATAC CTACAATATC
    901  TTGATTGCCA TCGGTTTGT TTAATTAC CAAAACGGGC TGACCTGCT
    951  TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
   1001  CTATGCACAT TATCATGTTT GCGTTGCAC TCATCCTGTT GCGCGTCCGC
   1051  AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
   1101  GAAAGCGCGA AAATGA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
     51  LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
    101  PVMQFAVPFA VLAVVMQLWV IPWAEALSRE YAEILKQKQE LSLVEAGEFN
    151  SLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF AKEGNFSLND
    201  NKRTLELRHG YRSGTPGRA DYNQVSFQKL NLIISTPKL IDPVSHRRTI
    251  PTAQLIGSSN PQHQAELMWR ISLTVSVLLL CLLAVPLSYF NPSRGHTYNI
    301  LIAIGLFLIY QNGLTLLFEA VEDGKIHFVL GLLPMHIIMF AVALILLRVR
    351  SMPSQPFWQA VGKSLTLKGG K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

5	orf101.pep	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVWX	10	20	30	40	50
	orf101a	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRX-AIDAVLALVGFVWXXM	10	20	30	40	50
10	orf101.pep	.....IAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL	90	100	110		
	orf101a	LTVSVLLLCLLAVPLSYFNPRSGHTYNILXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL	280	290	300	310	320
15	orf101.pep	LPMHIIMFVLALILLRVRSMPSQPFQAVGKSLTLKGGKX	120	130	140	150	
	orf101a	LPMHIIMFVIAIVLLRVRSMPSQPFQAVGKSLTLKGGKX	340	350	360	370	

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

1	ATGATTTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CGGCCGTCGG
51	CATTTTCGTC	GTCCTCTTGG	CGGTATTGGT	CTCCACGCAG	GCAATCAACC
101	TGCTCGCCCN	TGCCCGCGAC	NGGCGTNTCG	CCATCGATGC	CGTGTTGGCA
151	TTGGTCGGCT	TCTGGGTCNN	NNGNATGACG	CCGCTTTTGC	TNGTGTGAC
201	CGCATTATC	AGTACGTTGA	CCGTGTTGAC	CCGCTACTGG	CGNGACAGCG
251	AAATGTCGGT	CTGGNTATCC	TGCGGATTGG	CATTGAAACA	ATGGATACGC
301	CCGGTGATGC	AGTTTGCCGT	GCCGTTTGCC	GTTTGTGTTG	CCGTCAATGCA
351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTACG	CAGCCGCGAA	TACGCTGAAA
401	TCCTGAAGCA	GAAGCAGGAA	TGTCTTTGG	TGGAGGCAGG	CGGGTTCAAC
451	AGTTTGGGCA	AGCGCAACGG	CAGGGTTTAT	TTTGTGCGAA	CCTTCGATAC
501	CGAATCCGGC	ATCATGAAAA	ACCTGTTCCT	GCGCGAACAG	GACAAAAACG
551	GCGGCGACAA	CATCATCTTC	NCCAAAGAAA	GTAACCTCTC	GCTGAACGAC
601	AACAAACGCA	CGCTCGAATT	GCGCCACGGC	TACCGTTACA	GCGGCACGCC
651	CGGACGCGCC	GACTACAATC	AGGTTTCCTT	CCNAAAACCTC	AACCTGATTA
701	TCAGCACCAC	GCCCAAACTC	ATCGACCCCG	TTTCCACCCG	CCGTACNATN
751	CCNACNGCCC	AACTGATTGG	CAGCAGCAAC	CCGCAACATC	ANGCGGAATT
801	GATGTGGCGC	ATCTCGCTGA	CCGTGACGGT	CCTCCTACTC	TGCCTGCTTG
851	CCGTGCCGCT	TTCCTATTTT	AACCCGCGCA	GCGGACATAC	CTACAATATC
901	TTGANTGCCA	TCGGTTTGTT	TTTAATTAC	CAAAACGGGC	TGACCCGTGCT
951	TTTTGAAGCC	GTGGAAGACG	GCAAAATCCA	TTTTTGGCTC	GGACTGCTGC
1001	CTATGCACAT	CATCATGTTC	GTCATCGCAA	TCGTACTTCT	GCGCGTCGCG
1051	AGCATGCCCA	GCCAGCCCTT	CTGGCAGGCG	GTTGGCAAAA	GTCTGACATT
1101	GAAAGCGGGA	AAATGA			

45 This encodes a protein having amino acid sequence <SEQ ID 502>:

1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGXAAD	XXXAIDAVLA
51	LVGFWVXXMT	PLLLVLTAFT	STLTVLTRYW	RDSEMSVWXS	CGLALKQWIR
101	PVMQFAVPFA	VLVAVMQLWV	IPWAEILRSRE	YAEILKQKQE	LSLVEAGGFN
151	SLGKRNGRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIF	XKESNFSLND
201	NKRTLELRHG	YRYSGTPGRA	DYNQVSFXKL	NLIISTPCKL	IDPVSHRRTX
251	PTAQLIGSSN	PQHXAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
301	LXAIGLFLTY	QNGLTLLFEA	VEDGKIHEWL	GLLPMHIIMF	VIAIVLLRVR
351	SMPSQPFQQA	VGKSLTLKGG	K*		

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

55	orf101a.pep	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRXAIDAVLALVGFVWVXXMT	60
	orf101-1	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWVIGMT	60
60	orf101a.pep	PLLLVLTAFTISTLTVLTRYWRDSEMSVWXS	120
	orf101-1	PLLLVLTAFTISTLTVLTRYWRDSEMSVWLS	120

-297-

```

5      orf101a.pep      IPWAE LRSREYAEILKQKQELSLVEAGGFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ 180
      orf101-1          IPWAE LRSREYAEILKQKQELSLVEAGEFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ 180
      orf101a.pep      DKNGGDNIIFXKESNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFXKLNLIISTTPKL 240
      orf101-1          DKNGGDNIIFAKEGNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNLIISTTPKL 240
10     orf101a.pep      IDPVSHRRTXPTAQLIGSSNPQHXAELMWIRISLTVSVLLLC LLAVPLSYFNPRSGHTYNI 300
      orf101-1          IDPVSHRRTIPTAQLIGSSNPQHQAELMWIRISLTVSVLLLC LLAVPLSYFNPRSGHTYNI 300
15     orf101a.pep      LXAIGLFLIYQNGLTLLFEAVEDGKIHFWLG LLLPMHIIMFVIAIVLLRVRSMPSQPFWQA 360
      orf101-1          LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLG LLLPMHIIMFAVALILLRVRSMPSQPFWQA 360
      orf101a.pep      VGKSLTLKGGK      371
20     orf101-1          VGKSLTLKGGK      371

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.*

#### *gonorrhoeae*:

```

25     orf101.pep      MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFWV      57
      orf101ng         MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFWVIGM      59
30                                     //
      orf101.pep                                     IAIGLFLIYQNGLTLLFEAVEDGKIHFWLG      333
      orf101ng         SLTVSVLLLC LLAVPLSYFNPRSGHTYNI LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLG      331
35     orf101.pep      LLPMHIIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGK      373
      orf101ng         LLPMHIIIMFVIAIVLLRVRSMPSQPFWQAVG      362

```

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial amino acid sequence <SEQ ID 504>:

```

45     1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
      51  LVGFWVIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
      101  PVMQFAVPFA ILIAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
      151  NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF AKEGNFSLKD
      201  NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
      251  STAQLIGSSN PQHQAE LMRW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
      301  LIAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIIMF VIAIVLLRVR
      351  SMPSQPFWQA VG...

```

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

```

50     1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTGTGGT GTCCACGCAG GCGATCAACC
      101  TGCTTGGCCG CGCAGCTGAC GGGCGTGTG CCATCGATGC CGTGTGGCC
      151  TTAGTCGGCT TCTGGGTCAT CGGTATGACC CCGCTTTTGC TGGTGTGAC
      201  CGCATTATC AGCACGCTGA CCGTATTGAC CCGCTACTGG CGCGACAGCG
55     251  AAATGTCGGT CTGGCTATCC TCGGATTGG CGTTGAAACA GTGGATACGC
      301  CCCGTCATGC AGTTTGCCGT GCCGTTTGCC ATCCTGATTG CCGTCATGCA
      351  GCTTTGGGTG ATACCGTGGG CAGAGCTGCG CAGCCGCGAA TATGCCGAAA
      401  TTTTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAAGCCGG CGAGTTCAAT
      451  AACTTGGGCA AGCGCAACGG CAgggtttaT Ttcgtcgaaa CCTTTGACAC
60     501  CGaatccgGC ATCATGAAAA ACCTGTtct GcGCGAACAG GACAAAAACG
      551  gcggcgacaa CATCATCTTC GCaaaGAg gtaactTctc gctgaaggac

```

-298-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

601 AACAAAcgca cgctcgaATT GCGCCACGGC TACCGTTACA GCGGcacgcC
651 CGGacGCGCc gactaCAATC AGGTTtcctt cCAAAAcTc aacctgATta
701 TCAGCACACAC GCCCAAacTT ATCGaccCCG TTCCCAACCG CCGCACCAT
751 tcgacCGCCC AAcTGATTGG CAGCAGCAAT CCGCAACATC AGGCAGAAAT
801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTGCTC TGCCTACTCG
851 CCGTGCCGCT TTCCTATTTC AACCCGCGCA GCGGACATAC CTACAATATC
901 TTGATTGCCA TCGGTTTGT TTTAATTAC CAAAACGGGC TGACCCTGCT
951 TTTTGAAGCC GTGGAAGACG GCAAATCCA TTTTGGGCTC GGACTGCTGC
1001 CTATGCACAT CATCATGTTT GTCATCGCAA TCGTACTTCT GCGCGTCCGC
1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGgcgCA AAATGA

```

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
51 LVGFWVIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
101 PVMQFAVPFA ILIAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDN IIF AKEGNFSLKD
201 NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
251 STAQLIGSSN PQHQAE LMRW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
301 LIAIGLFLIY QNGLTLLFEA VEDGKIHF WL GLLPMHIIMF VIAIVLLRVR
351 SMPSQPFWQA VGKSLTLKGG K*

```

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60

```

      10      20      30      40      50      60
orf101-1.pep MIYQRNLIKELSF TAVGIFV VLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWVIGMT
      |||
orf101ng-1   MIYQRNLIKELSF TAVGIFV VLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWVIGMT
      |||

      70      80      90     100     110     120
orf101-1.pep PLLLVLTAFISTLT VLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV
      |||
orf101ng-1   PLLLVLTAFISTLT VLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAILIAVMQLWV
      |||

      130     140     150     160     170     180
orf101-1.pep IPWAE LRSREYAEILKQKQELSLVEAGEFN SLGKRNGRVYFVETFDTESGIMKNLFLREQ
      |||
orf101ng-1   IPWAE LRSREYAEILKQKQELSLVEAGEFN SLGKRNGRVYFVETFDTESGIMKNLFLREQ
      |||

      190     200     210     220     230     240
orf101-1.pep DKNGGDN IIFAKEGNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNLIISTTPKL
      |||
orf101ng-1   DKNGGDN IIFAKEGNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNLIISTTPKL
      |||

      250     260     270     280     290     300
orf101-1.pep IDPVSHRRTIPTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI
      |||
orf101ng-1   IDPVSHRRTISTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI
      |||

      310     320     330     340     350     360
orf101-1.pep LIAIGLFLIYQNGLTLLFEAVEDGKIHF WL GLLPMHIIMFAVALILLRVRSMPSQPFWQA
      |||
orf101ng-1   LIAIGLFLIYQNGLTLLFEAVEDGKIHF WL GLLPMHIIMFVIAIVLLRVRSMPSQPFWQA
      |||

      370
orf101-1.pep VGKSLTLKGGKX
      |||
orf101ng-1   VGKSLTLKGGKX
      |||

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is



The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

Computer analysis of this amino acid sequence gave the following results:

20 ORF and pspA show 44% aa identity in 179aa overlap:

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

50 1 MNKTLRYVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH  
51 SKAFCFGALG FSLCLALGTV NIAFADGIIT DKAAPKTOOA TILOTGNGIP

-300-

```

101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN
201 ASRATLTGQ PQYQAGDFSG FKIRQGNVAVI AGHGLDARDT DFTRILVCQQ
251 NHLDQYGRTS RHS*

```

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

10      1  ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
      51  CAACATTTCAT CTGGGTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTA
101    101  GCCATCATGC GCCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT
      151  TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
201    201  ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTGTA ACCGATCCAC
      251  GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GC+GGACAGC
15    301  CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
      351  CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
401    401  GTTTAGAcGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
      451  AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
501    501  AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGTTGGTAC
20    551  AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA
      601  CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
      651  GTTGTcAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
701    701  CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
      751  GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
25    801  ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
      851  TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
901    901  CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
      951  TATCACAGGC AAAGAAAAAG GTGTTT..

```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

30      1  ..STGHSEQNYT LPREITRNIS LGSFAYESHR KALSHHAPSQ GTELPQSNIGI
      51  SLPYTSNSFT PLPSSSLYII NPVNKGYLEV TDPRFANYRQ WLGS DYMLDS
101    101  LKLDPNNLHK RLGdGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
      151  NGATAARSMN LSVGIALSAE QVAQLTSDIV WLvQKEVKLP DGQTQTVLVP
35    201  QVYVRVKNd IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
      251  DNIGGRIHAQ KSAVTATQDI NNIGMLSAE QTLNLAGNN INSQSTTASS
301    301  QNTQGSSTYL DRMAGIYITG KEKGV..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and *pspA* protein show 50% aa identity in 325aa overlap:

```

40    Orf115: 1  STGHSEQNYTLPREITRNISLGSFAYESHRKALSHHAPSQGTLPQSNIGISLPYTSNSFT 60
      STG+S  Y  E++ +I +G AY+ +  +  P  +  NGI  +T
      pspA:  778 STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQQASDIPGTVPVVAENGIHPTFT----- 831
45    Orf115: 61  PLPSSSLYIINPVNKGYLEVTDPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQR 120
      LP+SSL+ I  P  NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQR+
      pspA:  832 -LPNSSLFAIAFPNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHIHKRLGDGYEQK 890
      Orf115: 121 LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
      L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV
50    pspA:  891 LVNEQIAKLTGYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIV 950
      Orf115: 181 WLvQKEVKLPDGGTQTVLVPQVYVRVKNdIDGKGALLSGSNTQINVSGSLKN-SGTIAG 239
      WL + V LPDG TQTVL P+VYVR + D++G+GALLSGS I SG+++N G IAG
      pspA:  951 WLENETVTLPDGTTQTVLKPVKYVRAREPKDMNGQGALLSGSVVDIG-SGAENRGGLIAG 1009
55

```

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from  
*N.gonorrhoeae*:

15	orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESH	31
	orf115.ng	NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDISLGSFAYESH	71
20	orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYL	81
	orf115.ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYL	131
25	orf115.pep	DPRFANYRQWLGS DYMLDSLKLD PNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND	141
	orf115.ng	DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND	191
30	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWL VQKEVKLPDGGTQTVLV	201
	orf115.ng	EEQFKALMDNGATAARSMNLSVGIALSAEQQAQLTSDIVWL VQKEVKLPDGGTQTVL	251
35	orf115.pep	VYVRVKNGDIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQ	261
	orf115.ng	VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQ	311
40	orf115.pep	SAVTATQDINNIGGMLSAEQTLNAGNINNSQSTTASSQNTQGSSTYLDRMAGIYITG	321
	orf115.ng	SAVTATQDINNIGGILSAEQTLNAGNINNSQSTAKSSQNAQGSSTYLDRMAGIYITG	371
45	orf115.pep	EKGV	325
	orf115.ng	EKGVLAQAQGDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRRFANYRQ	WLGSDDYMLGS
45	151	LKLDPPNHLK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLWQKEVKLP	DGGTQTVLMP
	251	QVYVRVKNGG	IDGKGALLSG	SNTQINVSGS	LKNSGTTAGR	NALINTDTL
	301	DNIGGRHAQ	KSAVTATQDI	NNIGGILSAE	QTLNLAGN	INNQSTAKS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIAGQ	ISNQSDGQQT
50	401	RLQAGRDNIL	DTVQTKGYQE	IHFADNHTI	RGSTNEVGSS	IQTKGDTVLL
	451	SGNNLNAKAA	EVGSAGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
	501	GNKLVIIDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDA NILGS	NVISDNGTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTQKSGLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTTVGS	LKGDTTIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSMIDGAAQ
55	651	NQLNSKTTQT	YEQKGLTVAF	SSPVTDLAQQ	AI AVAHKAAK	QFDKAKTTAL
	701	MPWRLPQVQ	RLFKQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

60

1	TTGCTTGTC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
51	CGAGAAAGAA	GTCTTCAGCG	AAATATGGTA	GTTCGACAA	TACTGGCGGT
101	CGCGTCGTAA	AGGACATGAT	GAAACAGGGC	ATCGTGAA	AAATTACT
151	TTGCCGGAGG	AAATCACACG	CGACATTTC	CTGGGTTT	TTGCCCTATG
201	ATCGCATAG	AAAGCATTAA	GCGGTCATGC	GCCGAGCC	GGCATCGAT
251	TGCCCAAAG	TAACCGGGAT	AATATCGTA	CTCGGAAAG	CAACGGTATT

-302-

301 TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT  
 351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC  
 401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC  
 451 CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA  
 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC  
 551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT  
 601 AATGGCGCGA CTGGCGCAGC TTCGATGAAT CTCAGCGTTG GCATTGCATT  
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATGTGT TGGTTGGTAC  
 701 AAAAAAGAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA  
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT  
 801 GTTGTGAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT  
 851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 901 CAAATATATG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC  
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT  
 1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT  
 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 1101 TATCAGAGGC AAAGAAAAAG GTGTTTATAG AGCGCAGGCA GGCAAGACA  
 1151 TCAACATCAT TGCCGCTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGCTGCAGG CAGGACGCGA CATTAACTTG GATACGGTAC AAACCGGCAA  
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA  
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGCGGCG  
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAAACC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAACCA ATCCCAAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT  
 1851 TGTTCGAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA  
 1951 AACCATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATTG AGTTGCCCCG TTACCGATTT GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAAC GACCGGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCTATCA AACAGGCAAA  
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40 1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPPEITRDIS LGSFAYESHK KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPREFANYRQ WLGS DYMLGS  
 151 LKLDPNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVOKEVKLP DGGTQTVLMP  
 45 251 QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 301 NNIGGILSAE QTLNLLNAGNN INNQSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINIAGQ ISNQSDQGQT  
 401 RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL  
 451 SGNLNAKAA EVGSAGKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 50 501 GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNTRI  
 551 QAGNHVRIGT TOTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ  
 651 NQLNSKTTQT YEQKGLTVAE SSPVTDLAQQ AIAVAHAHAN KSDKAKTTAL  
 701 MPWRLPMQVG RPIKQAKAHK T\*

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

20 30 40 50 60 70  
 orf115ng-1.p NEQTFGEKKVFSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDISLGSFAYESHK  
 orf115 STGHSEQNYTLPREITRNISLGSFAYESHK  
 10 20 30  
 80 90 100 110 120 130  
 orf115ng-1.p ALSRHAPSQGTGTELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET  
 orf115 ALSHHAPSQGTGTELPQSN-----GISLPYTPNSFTPLPSSSLYIINPVNKGKLVET  
 40 50 60 70 80

-303-

```

      140      150      160      170      180      190
orf115ng-1.p DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLGD GYEQRLNEQIAELTGHRRLDGYQND
      |||
orf115      DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLGD GYEQRLNEQIAELTGHRRLDGYQND
      90      100      110      120      130      140

      200      210      220      230      240      250
orf115ng-1.p EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLPDGGTQTVLMPQ
      |||
orf115      EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWL VQKEVKLPDGGTQTVLVPQ
      150      160      170      180      190      200

      260      270      280      290      300      310
orf115ng-1.p VYVRVKNGGIDGKGALLSGSNTQINVSGLKNSGTIAGRNALIINTDTLDNIGGRIHAQK
      |||
orf115      VYVRVKNGDIDGKGALLSGSNTQINVSGLKNSGTIAGRNALIINTDTLDNIGGRIHAQK
      210      220      230      240      250      260

      320      330      340      350      360      370
orf115ng-1.p SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK
      |||
orf115      SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK
      270      280      290      300      310      320

      380      390      400      410      420      430
orf115ng-1.p EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTKYQEIHFADNHTIR
      |||
orf115      EKG

```

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

```

30  gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length
    = 2273
    Score = 604 bits (1541), Expect = e-172
    Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

35  Query: 1   LLVQTEKDLHNEQTFGEKKVFSENGKLHNYWRARRKGHDETCHREQNYTLPEEITRDIS 60
    L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
    Sbjct: 739 LIVGTPESALDNDDELGTGTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR 796

40  Query: 61   LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
    +G AY+ + AP Q +++P + + NGI +T LP SSL+ I
    Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIIHPTFT-----LPNSSLFAI 840

45  Query: 121  NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLD PNNLHKRLGD GYEQRLNEQIAELT 180
    P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+L+NEQIA+LT
    Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSGYMLAALQODPNHIIHKRLGDGYEQKLVNEQIAKLT 900

50  Query: 181  GHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLP 240
    G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
    Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQITPGIALSAEQVARLTSDIVWLENETVTLP 960

55  Query: 241  DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGLKN-SGTIAGRNALIINTDT 299
    DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
    Sbjct: 961 DGTQTVLKPVKYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

60  Query: 300  LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTY 359
    + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
    Sbjct: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI- GAENALLKASNNIESRSETRSNQNEQGSVRN 1078

65  Query: 360  LDRMAGIYITGKEKGVLAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTKYQ 419
    + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q
    Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

70  Query: 420  EIHFDADNHTIRGSTNEVGSSIQTGKDVTLISGNNLNAKAAEVGSAGKTLAVYAKNDITI 479
    FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +
    Sbjct: 1139 NTIFSDSNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRKLKLAAGRDIKV 1198

    Query: 480  SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539
    +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
    Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

```

-304-

Query: 540 SNVISDNGTRI QAGNHVRIGTTQTQSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNR 1318

5 Query: 599 QSNEHTGSTVGS LKGD TTIVASKHYEQ TGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658  
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKH YTTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378

10 Query: 659 QTYEQKGLTVAFSSP VTD 676  
 Q YEQKG+TVA S PV +  
 Sbjct: 1379 QVYEQKGVTV AISVPV V 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG  
 51 TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA  
 101 ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT  
 151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC  
 201 CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 251 ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT  
 301 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG  
 351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG  
 401 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC  
 25 451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT  
 501 TGTTCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG  
 551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC  
 601 AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAA  
 651 GGTGGCATTC AGTTCGCCCG TTACCGATTT GGCACAACAA ...

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

1 ..SGNNLNAKAA EVSSANGTLA VSANNDINIS AGINTHVDD ASKHTGRSGG  
 51 GNKLVIDKA QSHHETAQSS TFEKQVVLQ AGNDANILGS NVISDNGTQI  
 101 QAGNHVRIGT TQTQSQSETY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS  
 151 NEHTGSTVGS LKGD TTIVAG KHYEQIGSTV SSPEGNN TIY AQSIDIQAAH  
 35 201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

40 Orf117: 4 NLNAKAAEVSSANGTLAVSANNDINISAGINTHVDDASKHTGRSGG GNKLVIDKAQSH 63  
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++  
 pspA: 1173 DIRIRAAEVGSEQGR LKLAAGRDIKVEAGKAHTETEDALKYTGRSGGGIKQKMTRHLKNQ 1232

45 Orf117: 64 HETAQSSTFEKQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSETYHQT 123  
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++  
 pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTG SNIIADNHTILSAKNNIVLKAETRSRSAEMNK 1292

50 Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGS LKGD TTIVAGKHYEQIGSTVSS 182  
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGS L G+T I AGKHY Q GST+SS  
 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKH YTTQTGSTISS 1352

Orf117: 183 PEGNN TIYAQSIDIQAHNKLNSNTTQTYEQKXLTVAFSSP VTD 226  
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +  
 pspA: 1353 PQGDVGISSGKISIDAAQNRYSQESKQVYEQKGVTV AISVPV V 1396

Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

5	orf117.pep	SGNNLNAAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFDADNHTIRGSTNEVGSSIQTKGDVTLTSGNNLNAAAEVSGAKGTLAVYAKNDITIS	480
10	orf117.pep	AGINTTHVDDASKHTGRSGGKNKLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
	orf117ng	SGIHAGQVDDASKHTGRSGGKNKLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
15	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSEYHQTQKSLMSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	NVISDNGTQIQAGNHVRIGTTQTQSQSEYHQTQKSLMSAGIGFTIGSKTNTQENQSQS	600
20	orf117.pep	NEHTGSTVSLKGDTTIVAGKHYEQIGSTVSSPEGNNIIYAQSIDIQAHNKLSNNTTQT	210
	orf117ng	NEHTGSTVSLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTTQT	660
25	orf117.pep	YEQKXLTVAFSSPVTDLAQQ	230
	orf117ng	YEQKGLTVAFSSPVTDLAQQAIAVAHKAQKQFDKAKTTALMPWRLPMQVGRLEFKQAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
	51	LPEEITRDIS LGSFAYESH KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
30	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPREFANYRQ WLGS DYMLGS
	151	LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
35	201	NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
	251	QVYVRVKNKG IDGKALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
40	301	DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLLLNAGNN INNQSTAKSS
	351	QNAQGSSTYL DRMAGIYITG KEGVLAQA GKDINIIAGQ ISNQSDQGGT
45	401	RLQAGRDLNL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLT
	451	SGNNLNAAAEVSGAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
50	501	GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTTRI
	551	QAGNHVRIGT TQTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS
55	601	NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
	651	NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAQ QFDKAKTTAL
60	701	MPWRLPMQVG RLFQAKAPK K*

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

45	1	TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGACAAAC TACTGGCGTG
50	101	CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATTATACT
	151	TTGCCGGAGG AAATCACACG CGACATTCA CTGGGTTCAT TTGCCTATGA
55	201	ATCGCATAGC AAAGCATTAA GCCGTCATGC GCCAGCCAA GGCACGAGT
	251	TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT
60	301	TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
	351	ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGTA ACCGATCCAC
65	401	GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
	451	CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
70	501	CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGCATCGTC
	551	GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGCAT
75	601	AATGGCGCGA CTGCGGCAGC TTCGATGAAT CTCAGCGTTG GCATTGCATT
	651	AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATGTGT TGGTTGGTAC
80	701	AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
	751	CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
85	801	GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAACT
	851	CAGGCACGAT TGCGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
90	901	GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
	951	ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
95	1001	TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
	1051	CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA

1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA  
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAG  
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA  
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAC ATACAGGCAG AAGCGCGCGC  
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGAGC GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT  
 1851 TGTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA  
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATTC AGTTCGCCCC TTACCGATTT GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAC GACCGCGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPEEITRDIS LGSFAYESH S KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDRFANYRQ WLGS DYM LGS  
 151 LKLDPNNLHK RLGDGYEQ R LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP  
 251 QVYVRVKNNG IDKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGN NINQSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIIAGQ ISNQSDQGQT  
 401 RLQAGRDINL DTVQTKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL  
 451 SGNNLNAAK EVGSAKGT LA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 501 GNKLVIDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMIDGAAQ  
 651 NQLNSKTTQT YEQKGLTVA SFSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL  
 701 MPWRLPMQVG RPIKQAKAHK T\*

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it  
 shows homology with a secreted *N.meningitidis* protein in the database:

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length = 2273  
 Score = 604 bits (1541), Expect = e-172  
 Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)  
 Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDIS 60  
 L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I  
 Sbjct: 739 LIVGTPESALDNDDELGTCTI-TDKGDLHRYHRHHKGRDSTGYSRSPYEFAPPEVS-SIR 796  
 Query: 61 LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120  
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I  
 Sbjct: 797 MGISAYKGY-----APQASDIPGTV---VPVVAENGTHPTFT-----LPNSSLFAI 840  
 Query: 121 NPANKGYLVE TDRFANYRQWLGS DYM LGS LKLDPNNLHKRLGDGYEQRLINEQIAELT 180  
 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+L+NEQIA+LT  
 Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSYMLAALQQDPNHIHKRLGDGYEQKLVLNEQIAKLT 900  
 Query: 181 GHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLP 240  
 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTS DIVWL + V LP  
 Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTGIALSAEQVARLTS DIVWLENETVTL 960  
 Query: 241 DGGTQTVLMPQVYVRVKNNGIDKGALLSGSNTQINVSGSLKN-SGTIAGR NALIINTDT 299  
 DG TQTVL P+VYVR + ++G+ALLSGS I SG+++N G IAGR ALI+N  
 Sbjct: 961 DGTQTVLKP KPVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALIQAQN 1019  
 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGN NINQSTAKSSQNAQGSSTY 359



-307-

+ N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS  
 Sbjct: 1020 IKNLQGDLOGKNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078  
 Query: 360 LDRMAGIYITGKEKGVLAQAAGKDINIIAGQISNQSDQGQTRLOAGRDINLDTVQTGKYQ 419  
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q  
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138  
 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLISGNNLNAKAAEVGSAKGTILAVYAKNDITI 479  
 FD+DN+ IR NEVGSI+T+G+++L + ++ +AAEVGS +G L + A DI +  
 Sbjct: 1139 NTIFSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIKV 1198  
 Query: 480 SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539  
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G  
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258  
 Query: 540 SNVISDNTRIQAAGNHVRIGTTQTQSOSQETIYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNR 1318  
 Query: 599 QSNEHTGSTVGLKGDTTIVASKHYEQTSNVSSPEGNNLISTQSMIDIGAAQNQLNSKTT 658  
 ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378  
 Query: 659 QTYEQKGLTVAFFSSPVD 676  
 Q YEQKG+TVA S PV +  
 Sbjct: 1379 QVYEQKGVTVAISVPV 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCGCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTTCG  
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCACTCATG ATGCCGAAAC CCCAACCGGC  
 201 GGTCAAAAAA ACGGCAAAC CCAGAACCC CGCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAA AGCCTCCCGC  
 301 TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGCA TTATCGGCAA  
 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC  
 401 CTCCGACGC GTCGGCAAAC CCTGCACCG TTCCGCAAAC ACCTGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA  
 501 CGTGCGCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKVR DQFGHSDKDA LLNSXTSHVR  
 51 DGKPSGGSV MPKQPAVKK TAKPDPMXR NLQEQDAVYI AKQKQAKASP  
 101 FKTEIETALE ESGIIGNSAH TVSEFQTGHS ATKPADASAK PAPVPQTPAK  
 151 PLITLKEISK VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCGCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTTCG  
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCACTCATG ATGCCGAAAC CCCAACCGGC  
 201 GGTCAAAAAA ACGGCAAAC CCAGAACCC CGCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAA AGCCTCCCGC  
 301 TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGCA TTATCGGCAA  
 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC  
 401 CTCCGACGC GTCGGCAAAC CCTGCACCG TTCCGCAAAC ACCTGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA  
 501 CGTGCGCTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC  
 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTGCGC  
 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG

-308-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

651 CTATCAGGCA TTTATCGTGG GTATTCAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTTCGA
751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG
851 CCATCCATT TGGTTCCCGG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TGCCTGTT CTCCTAA

```

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15  
20  
25  
30  
35  
40

```

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGSSVM MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK
151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGFVLE DDGAFHYTDT SGSTMFSICS LNNPFTNAL LDNQSYPKGF
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVVLARQE MLKVGIEPGG KTAIRLFS*

```

Computer analysis of this amino acid sequence gave the following results:

#### 25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N. meningitidis*:

30  
35  
40  
45

```

          10      20      30      40      50      60
orf119.pep MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV RDGKPSGGSSVM
          |||||:|||||
orf119a    MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV RDGKPSGGPVM
          10      20      30      40      50      60

          70      80      90      100     110     120
orf119.pep MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
          |||||:|||||
orf119a    MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
          70      80      90      100     110     120

          130     140     150     160     170
orf119.pep TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRIDFISY
          |||||:|||||
orf119a    TVPEPQTGHSAPKPADAPAKPVVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE
          130     140     150     160     170     180

          190     200     210     220     230     240
orf119a    AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS

```

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

50  
55  
60

```

1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCTCGCCG TTGTGCCTA
51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTTCG
101 GGCACTCCGA CAAAGATGCC CTGCTCAACA GCAAACCAG CCATGTCCGC
151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCAACCGGC
201 GGTCAAAAAA ACGGCAAAAT CCCAAGACCC CGCCATGCGC AACCTGCAAG
251 AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGC
351 CTCCGCCAC ACCGTTCCCG AACCCCAAAC CGGACATTCC GCACCAAAAC
401 CTGCCGACGC GCCGGCAAAA CCTGTCCCG TTCCGCAAAAC GCCGGCAAAA
451 CCGCTGATTA CGTCAAAGA GCTGTCGAAG GTCGAGCTGC CCTGGTTGA
501 CGTGCGCTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAAGAAC
551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCGGC
601 TGCACCATGG ACGACATTT CCAGATTGCC GAACCCATCC CGGGCATCCG

```

-309-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

651 CTATCAGGCA TTTATCGTGG GTATTTCAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTGCGA
751 CACAGCATGG GCGGTGAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGCTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TCGCGCTGTT CTCCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 528>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
151 PLITLKELSK VELPWFDFVR DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVRLARQSE MLKVGIEPPG KTALRLFS*

```

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60  
65

```

orff119a.pep      10      20      30      40      50      60
MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV
RDGKPSGGPVM
orff119-1          10      20      30      40      50      60
MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV
RDGKPSGGPVM

orff119a.pep      70      80      90      100     110     120
MPKPQPAVKKTA
KSQDPAMRN
LQEQDAVYI
AKQKQAKAS
PFKTEIETA
LES
SGIIGNSAH
orff119-1          70      80      90      100     110     120
MPKPQPAVKKTA
KSQDPAMRN
LQEQDAVYI
AKQKQAKAS
PFKTEIETA
LES
SGIIGNSAH

orff119a.pep      130     140     150     160     170     180
TVPEPQTGHSAPK
PADAPAKPV
PVPQTPAKPLI
TLKELSKVEL
PWFDFVF
DFISYIALTE
orff119-1          130     140     150     160     170     180
TVSEPTGHSAPK
PADAPAKPV
PVPQTPAKPLI
TLKELSKVEL
PWFDFVF
DFISYIALTE

orff119a.pep      190     200     210     220     230     240
AKELHALPRLSN
RCRYQIVGCT
MDDHFQIAE
PIPGIRYQA
FIVGIQAVS
RNGLASQEELS
orff119-1          190     200     210     220     230     240
AKELHALPRLSN
RCRYQIVGCT
MDDHFQIAE
PIPGIRYQA
FIVGIQAVS
RNGLASQEELS

orff119a.pep      250     260     270     280     290     300
AFNRQVDAFAH
SMGGQTLHT
DLAAFIEVA
SALDAFCAR
VDQTIAIHL
VSP
TSISGVELRS
orff119-1          250     260     270     280     290     300
AFNRQVDAFAH
SMGGQTLHT
DLAAFIEVA
SALDAFCAR
VDQTIAIHL
VSP
TSISGVELRS

orff119a.pep      310     320     330     340     350     360
AVTGVGVFLEDD
GAFHYTDT
SGSTMFSIC
SLNNEPFTN
ALLDNQSYK
GFS
MLLDIPHSPA
orff119-1          310     320     330     340     350     360
AVTGVGVFLEDD
GAFHYTDT
SGSTMFSIC
SLNNEPFTN
ALLDNQSYK
GFS
MLLDIPHSPA

orff119a.pep      370     380     390     400     410     420
GEKTFDDLFDL
AVRLSGQLN
LNLVNDKME
EVSTQWLKDV
RTYVRLARQ
SEM
LKVGIEPPG
orff119-1          370     380     390     400     410     420
GEKTFDDLFDL
AVRLSGQLN
LNLVNDKME
EVSTQWLKDV
RTYVRLARQ
SEM
LKVGIEPPG

```

-310-

```

orf119a.pep    KTALRLFSX
               |||||
orf119-1       KTALRLFSX

```

### 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

```

10 orf119.pep    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRD GKPSGGGSVM    60
    orf119ng     MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGGPVM    60

15 orf119.pep    MPKPQPAVKKTAKPQDPXMRNLQE QDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH    120
    orf119ng     MPKPQPAVKKPAKPQDSAMRNLQE QDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH    120

    orf119.pep    TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLTKELSKVELSWFDFVRIDFISY    175
    orf119ng     TVSEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLTKELSKVELPWFDFVRDFISYIALTE    180

```

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

```

20      1  ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCTCGCCG TTGTGCGCTA
      51  CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
     101  GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAACACAG CCATGTCCGC
     151  GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACCGGC
     201  GGTCAAAAAA CCGGCCAAAC CCCAAGACTC CGCCATGCGC AACCTGCAAG
     251  AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
     301  TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAATCGGCA TTATCGGCAA
     351  CTCGCCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC
     401  CTGCCGACGC GCGGCAAAA CCGTTCCCG TTCCGCAAA GCGGGCAAAA
     451  CCGCTGATTA CGCTCAAAGA GCTGTCGAAG GTCGAGCTGC CCTGGTTTGA
     501  CGTGCCTtc gACTTCATCT CCTATATCGC GCTGACCGAA GCCAAGAAGC
     551  TGCACGCACT GCGCGCCTT tccAACCGCT GCGCTACCA GATTGTCGGC
     601  TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG
     651  CTATCAGGCA TTTATCGTGG GTATCCAGGC AGTCAGCCGC AACGGACTTG
     701  CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGCGGA CGCATTGCGA
     751  CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG COTTTATCGA
     801  AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG
     851  CCATCCATTT GGTTCGCGC ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
     901  GCGGTAACGG GCGTGGGTTT CTTTTTGGA GACGACGCG CGTTCCACTA
     951  TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
    1001  AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
    1051  ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGCAAAAAA CCTTCGACGA
    1101  TTTGTTTATG GATTGGCGG TACGCTGTC CGGTCAATTG AACCTGAATC
    1151  TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCAATGGCT CAAAGACGTA
    1201  CGCACTTATG TATTGGCGG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
    1251  ACCGGGCGGC AAAACCGCCC TGCGCCTGTT TTCATAA

```

This encodes a protein having amino acid sequence <SEQ ID 530>:

```

50      1  MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
     51  DGKPSGGPVM MPKPQPAVKK PAKPQDSAMR NLQE QDAVYI AKQKQAKASP
    101  FKTEIETALE EIGIIGNSAH TVSEPQTGHS APKPADAPAK PVPVPQTPAK
    151  PLITLTKELSK VELPWFDFRF DFISYIALTE AKELHALPRL SNRCRYQIVG
    201  CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQADAFa
    251  QSMGGQTLHT DLAAFIEVAS ALDAFCARVD Q'IAIHLVSP TSISGVELRS
    301  AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
    351  MLLDIPHSFA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
    401  RTYVLARQSE MLKVGIEPGG KTALRLFS*

```

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

```

60      10      20      30      40      50      60
    orf119ng    MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGPVM
    orf119-1    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGGSVM
                10      20      30      40      50      60

```

-311-

		70	80	90	100	110	120
	orf119ng	MPKPQPAVKKPAKPQDSAMRN	LQEQDAVYIAKQKQAKASPF	KTEIETALEEIGIIGNSAH			
5	orf119-1	MPKPQPAVKKTAKPQDPAMRN	LQEQDAVYIAKQKQAKASPF	KTEIETALEESGIIGNSAH			
		70	80	90	100	110	120
	orf119ng	TVSEPQTGHSAPKPADAPAKP	VPVPQTPAKPLITLKELSKVEL	PWFDVRFDFISYIALTE			
10	orf119-1	TVSEPQTGHSAPKPADAPAKP	APVPQTPAKPLITLKELSKVEL	PWFDVRFDFISYIALTE			
		130	140	150	160	170	180
	orf119ng	AKELHALPRLSNRCRYQIVGCT	MDDHFQIAEPIPIGIRYQAFIV	GIQAVSRNGLASQEELS			
15	orf119-1	AKELHALPRLSNRCRYQIVGCT	MDDHFQIAEPIPIGIRYQAFIV	GIQAVSRNGLASQEELS			
		190	200	210	220	230	240
	orf119ng	AFNRQADAFQSMGGQTLHTDL	AAFIEVASALDAFCARVDQTI	AIHLVSPTSISGVELRS			
20	orf119-1	AFNRQVDAFAQSMGGQTLHTDL	AAFIEVASALDAFCARVDQTI	AIHLVSPTSISGVELRS			
		250	260	270	280	290	300
	orf119ng	AVTGVGfVLEDDGAFHYTDTSG	STMFSICSLNNEPFTNALLDN	QSYKGfSMMLLDIPHSPA			
25	orf119-1	AVTGVGfVLEDDGAFHYTDTSG	STMFSICSLNNEPFTNALLDN	QSYKGfSMMLLDIPHSPA			
		310	320	330	340	350	360
	orf119ng	GEKTFDDLfMDLAVRLSGQLNL	NLVNDKMEEVSTQWLKDVRTY	VLARQSEMLKVGIIEPGG			
30	orf119-1	GEKTFDDLfMDLAVRLSGQLNL	NLVNDKMEEVSTQWLKDVRTY	VLARQSEMLKVGIIEPGG			
		370	380	390	400	410	420
	orf119ng	KTALRLFSX					
35	orf119-1	KTALRLFSX					
		429					
40	orf119ng	KTALRLFSX					
	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	..GCGCGGCACG	GCACGGAAGA	TTTCTTCATG	AACAACAGCG	ACAC.ATCAG
	51	GCAGATAGTC	GAAAGCACCA	CCGCTACGAT	GAAGCTGCTG	ATTTCCTCCA
50	101	TCGCCCTGAT	TTCATTGGTA	GTCGGCGGCA	TCGGCGTGAT	GAACATCATG
	151	CTGGTGTCG	TTACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGGCAAT
	201	CGGCGCGCGG	CGCGCAATA	TTTyGCAGCA	GTTTTTGATT	GAGGCGGTGT
	251	TAATCTGCGT	CATCGGCGGT	TTGGTCGGCG	TGGGTTTGTC	CGCCGCCGTC
	301	AGCCTCGTGT	TCAATCATT	TGTAACCGAC	TTCCCGATGG	ACATTTCGCG
	351	CATGTCCGTC	ATCGGCGCGG	TCGCCTGTTC	GACCGGAATC	GGCATCGCGT
55	401	TCGGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA
	451	TTGGCACAGG	ATTGA			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	..ARHGTEDEFFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIM
	51	LVSVTERTKE	IGIRMAIGAR	RGNIXQQFLI	EAVLICVIGG	LVGVGLSAAV
60	101	SLVFNHFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGEMPAN	KAALKNPIDA
	151	LAQD*				

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

```

      1 ATGTCGGTGC AAGCAGTATT GCGGCACAAA ATGCGTTCGC TTCTGACGAT
      51 GCTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTCCTGTC GTCGCATTGG
    101 GCAATGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
  5   151 AACACCATCA GCATCTTCCC GGGGCGCGGC TTCGCGGACA GGCGCAGCGG
      201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
      251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
      301 TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACAATA
    10  351 TTTCGACGTG CGCGGACTGA AGCTGGAAAC GGGGCGGCTG TTTGACGAAA
      401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
      451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTTCAG
      501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAAACGCTT
      551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGCTGATG
    15  601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
      651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
      701 AAGCGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
      751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCT
      801 CATCGCCCTG ATTTTCATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
    20  851 TGCTGGTGTC CGTTACCGAG CGCACCAGG AAATCGGCAT ACGGATGGCA
      901 ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTGTGA TTGAGGCGGT
      951 GTTAATCTGC GTCATCGGCG GTTTGGTCGG CETGGGTTTG TCCGCGCGCG
    25 1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC
      1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
      1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGACG
      1151 CATTGGCACA GGATTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

```

      1 MSVQAVLAHK MRSLLTMLGI IIGIASVVSV VALNGSQKK ILEDISSIGT
      51 NTISIFPGRG FGDRRSRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
    30 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
      151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
      201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTG DFFMNSDSI
      251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
      301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
      351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E.coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

```

Orf134: 2  RHGTEDFFMNSDXIRQIVESTTGTMKXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEI 61
      RHG +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EI
    40 o648: 496 RHGKKDFFTWNMDGVLTVEKTRTLQLFLTLVAVISLVVGIGVMNIMLVSVTERTREI 555
      Orf134: 62  GIRMAIGARRGNIXQFLIEAXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAMSVI 121
      GIRMA+GAR ++ QQFLIEA F+ + + S +++++
    45 o648: 556 GIRMAVGARASDVLQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPALAL 615
      Orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD 154
      A CST GI FG++PA AA+L+P+DALA++
    o648: 616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648
  
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N. meningitidis*:

```

      orf134.pep
    55 orf134a  GESHTNSITVKIKDNANTQVAEKGLTDLKARHGTEDFFMNSDSIRQIVESTTGTMKLL
      210      220      230      240      250      260
      40      50      60      70      80      90
  
```

-313-

```

5  orf134.pep  ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
   orf134a    ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG
                                     270      280      290      300      310      320

10  orf134.pep  LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
   orf134a    LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
                                     330      340      350      360      370      380

15  orf134.pep  LAQDX
   orf134a    LAQDX

```

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

```

1  ATGTCGGTGC AAGCAGTATT GCGGCACAAA ATGCGTTCGC TTCTGACGAT
51  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCCGATTGG
101 GCAACGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
151 AACACCATCA GCATCTTCCC AGGGCGCGGC TTCGGCGACA GGCGCAGCGG
201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
351 TTTCGACGTG CGCGGGCTGA AGCTGGAAC GGGCGGCTG TTTGACGAAA
25  401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAACCA TTTGTTCAG
501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
30  651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
701 AAGCGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCTC
801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
35  851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGCGC
901 ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
951 GTTAATCTGC GTCATCGGCG GTTTGGTCGG CGTGGGTTTG TCCGCCGCCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC
1051 GGCATGTCCG TCATCGGCGC GGTCCGCTGT TCGACCGGAA TCGGCATCGC
40  1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGATG
1151 CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 536>:

```

1  MSVQAVLAHK MRSLLTMLGI IIGIASVVSVALGNGSQKK ILEDISSIGT
51  NTISIFPGRG FGDRRSGRIG TLTIDDAKII AKQSYVASAT PMTSSGGTTLT
45  101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
151 DKLFDSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

50 ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

```

55  orf134a.pep  MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG
   orf134-1     MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG

60  orf134a.pep  FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
   orf134-1     FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

65  orf134a.pep  RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1     RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

   orf134a.pep  ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
   orf134-1     ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE

```

-314-

```

5  orf134a.pep  DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134-1    DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134a.pep  IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
   orf134-1     IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
10  orf134a.pep  STGIGIAFGFMPANKAAKLNPIDALAQDX
   orf134-1     STGIGIAFGFMPANKAAKLNPIDALAQDX

```

Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

```

20  orf134.pep          ARHGTE DFFMNSDXIRQIVESTTGTMKLL  30
   orf134ng            GESHTNSITVKIKDNANTRVAEKGLAELLKARHGTE DFFMNSDSIRQIVESTTGTMKLL  264
   orf134.pep          ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG  90
   orf134ng            ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIGG  324
25  orf134.pep          LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA  150
   orf134ng            LVGVGLSAAVSLVFNHFVTDFFPMDISAAVIGAVACSTGIGIAFGFMPANKAAKLNPIDA  384
30  orf134.pep          LAQD  154
   orf134ng            LAQD  388

```

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

```

35  1  ATGTCGGTGC  AAGCAGTATT  GGCACACAAA  ATGCGTTCGC  TTCTGACCAT
   51  GCTCGGCATC  ATCATCGGTA  TCGCTTCGGT  TGTCTCCGTC  GTCGCGCTTG
  101  GCAACGGTTC  GCAGAAAAAA  ATCCTCGAAG  ACATCAGTTC  GATGGGGACG
  151  AACACCATCA  GCATCTTCCC  CGGGCGCGGC  TTCGCGCACA  GGCGCAGCGG
  201  CAAAATCAAA  ACCCTGACCA  TAGACGACGC  AAAAATCATC  GCCAAACAAA
  251  GCTACGTTGC  CTCGCCACG  CCCATGACTT  CGAGCGGCGG  CACGCTGACC
  301  TACCGCAATA  CCGACCTGAC  CGCTTCTTTG  TACGGTGTGG  GCGAACAATA
  351  TTTTCGACGTG  CGCGGGCTGA  AGCTGGAAAC  GGGGCGGCTG  TTTGATGAGA
  401  ACGATGTGAA  AGAAGACGCG  CAAGTCGTCG  TCATCGACCA  AAATGTCAAA
  451  GACAAACTCT  TTGCGGACTC  GGATCCGTTG  GGTAAAACCA  TTTTGTTTCA
  501  GAAACGCCCC  TTGACCGTCA  TCGGCGTGAT  GAAAAAGAC  GAAAACGCTT
  551  TCGGCAATTC  CGACGTGCTG  ATGCTTTGGT  CGCCCTATAC  GACGCTGATG
  601  CACCAATCA  CAGGCGAGAG  CCACACCAAC  TCCATCACCG  TCAAAATCAA
  651  AGACAATGCC  AATACCCGGG  TTGCCGAAAA  AGGGCTGGCC  GAGCTGCTCA
  701  AAGCACGGCA  CGGCACGGAA  GACTTCTTTA  TGAACAACAG  CGACAGCATC
  751  AGGCAGATGG  TCGAAAGCAC  CACCGGTACG  ATGAAGCTGC  TGATTTCCTC
  801  CATCGCCCTG  ATTTTCATTG  TAGTCGGCGG  CATCGGTGTG  ATGAACATTA
  851  TGCTGGTGTC  CGTTACCGAG  CGCACCAAAG  AAATCGGCAT  ACGGATGGCA
  901  ATCGGCGCGC  GGCGCGGCAA  TATTTTGACG  CAGTTTTTGA  TTGAGGCGGT
  951  GTTAATCTGC  ATCATCGGAG  GCTTGGTTCG  CGTAGGTTTG  TCCGCCGCCG
1001  TTAGCCTCGT  GTTCAATCAT  TTTGTAACCG  ATTTCCCGAT  GGACATTTTC
1051  GCGGCATCCG  TTATCGGGGC  GGTGCGCTGT  TCGACCGGAA  TCGGCATCGC
1101  GTTCGCTTT  ATGCTGCCA  ATAAGGCAGC  CAAACTCAAT  CCGATAGATG
1151  CATTGGCGCA  GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 538>:

```

60  1  MSVQAVLAHK  MRSLLTMLGI  IIGIASVSV  VALNGSQKK  ILEDISSMGT
   51  NTISIFPGRG  FGDRRSQKIK  TLTIDDAKII  AKQSYVASAT  PMTSSGGTLT
  101  YRNTDLTASL  YGVGEQYFDV  RGLKLETGRL  FDENDVKEDA  QVVVIDQNVK
  151  DKLFADSDPL  GKTILFRKRP  LTVIGVMKKD  ENAFGNSDVL  MLWSPYTTVM
  201  HQITGESHTN  SITVKIKDNA  NTRVAEKGLA  ELLKARHGTE  DFFMNSDSI
  251  RQMVESTTGT  MKLLISSIAL  ISLVVGGIGV  MNIMLVSVTE  RTKEIGIRMA
  301  IGARRGNILQ  QFLIEAVLIC  IIGGLVGVGL  SAAVSLVFNH  FVTDFPMDIS

```



351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD\*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5   orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSMGNTNTISIFPGRG
    orf134-1      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG

10  orf134ng      FGDRRSQGIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
    orf134-1      FGDRRSQGIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

15  orf134ng      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLFRKRPLTVIGVMKKD
    orf134-1      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLFRKRPLTVIGVMKKD

20  orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGTE
    orf134-1      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE

25  orf134ng      DFFMNNSDSIRQMVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
    orf134-1      DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA

    orf134ng      IGARRGNILQQFLIEAVLICIIIGGLVGVGLSAAVSLVFNHFVTDPFMDISAASVIGAVAC
    orf134-1      IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDPFMDISAMSVIGAVAC

    orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
    orf134-1      STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E.coli* ABC transporter:

```

35  sp|P75831|YBJZ_ECOLI_HYPOTHETICAL_ABC_TRANSPORTER_ATP-BINDING_PROTEIN_YBJZ >gi5
    (AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length =
    648
    Score = 297 bits (753), Expect = 6e-80
    Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

40  Query: 1      MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXGNGSQKKILEDISSMGNTNTISIFPGRG 60
    M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
    Sbjct: 260 MAWRALAANKMRTLMLGIIIGIASVVSIVVVGDAAKQMVLAIRSIGTNTIDVYPGKD 319

45  Query: 61     FGDRRSQGIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
    FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
    Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

50  Query: 121    RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTLFRKRPLTVIGVMKK 179
    G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
    Sbjct: 380 YGMTFSEGNTFNQEQLNGRAQVVVLDSNTRRQLFPHKADVGEVILVGNMPARVIGVAEE 439

55  Query: 180    DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239
    ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
    Sbjct: 440 KQSMFGSSKVLRVWLPYSTMSGRVMGQSWLNSITVRVKEGFDSAEAEQQLTRLLSLRHGK 499

60  Query: 240    EDFFMNNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEIGIRM 299
    +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EIGIRM
    Sbjct: 500 KDFFTWNMDGVLTVEKTTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREIGIRM 559

    Query: 300    AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDPFMDISAASVIGAVA 359
    A+GAR ++LQQFLIE F+ + + S +++ A
    Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAFL 619

    Query: 360    CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
    CST GI FG++PA AA+L+P+DALA++
    Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1  ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCTTT
51  GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCTTCCTT GATTTTGAAA GAACGGATTT CCGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGC CGCATGT
251 CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACCC
301 GGCTGGCGCG TCGTGTTTTA CCTTTCCGTG ACAGGTGTGG CGATGTCGTC
351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCTTTT CCATCGGCAG
401 TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
15 451 ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTTCCTA
501 TATGACCGTC GTTTTTCCTG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CQTACGCGGT
601 ATTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

20 1  ..GTGAMLLLFY AVTILPLATG VTLSTSSIF LAVSFILILK ERISVYTQAV
51  LLLGFAGVVL LLNPSFRSGQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVWA TLTGWHTLSF PSAVYLSCIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELFWQEI LGMCIILISAV
201 F*

```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GGCGGCCTGC TTACCATTAT TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTGTCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC
30 201 GCCCCATTGG AAAAACCCTT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTTACCGG GTAACGCATC TGCCCTTTGGC CACTGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCTGAT
351 TTTGAAAGAA CGGATTTCGG TTTACACGCA GGCGGTGCTG CTCTCTGGTT
35 401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCGCAG CGGTCAGGAA
451 ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAAGTGT CTTTGGCGGG CGAACC CGG TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGCGCA TGTCGTGCGT TTGGGCGACG
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCTGT
651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
40 701 AAGTCGGCGA CAAATTCACG GTTGCTCGC TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCGC ATTTTTTCTG GGCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCTTCAAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMIFS
51  TVALGAAAVL RRDxFRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGV
101 TLSYTSSIFL AVSFILILKE RISVYTQAVL LLGFAGVVL LLNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVMSSVWAT
50 201 LTGWHTLSFP SAVYLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL GEELFWQEIL GMCIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N. meningitidis*:

```

5      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSTSSIF
                        |||
      orf135a      50      60      70      80      90      100
      STVALGAAAVLRRDTERTPHWKNHLNRSMTGTGAMLLLFYAVTHLPLATGVTLSTSSIF

10     orf135.pep      40      50      60      70      80      90
      LAVFSFLILKERISVYTQAVLLLGFGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
      |||
      orf135a      110     120     130     140     150     160
      LAVFSFLILKERISVYTQAVLLLGFGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK

15     orf135.pep      100     110     120     130     140     150
      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSICIGVSALIAQLSM
      |||
      orf135a      170     180     190     200     210     220
      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSICIGVSALIAQLSM

20     orf135.pep      160     170     180     190     200
      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVFX
      |||
      orf135a      230     240     250     260     270     280
      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLAEEELFWQEILGMCIIISGILSSIRPTAF

      orf135a      KQRLQSLFRQRX
      290      300

```

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GCGCGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTGTCCCT CCGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTGCGC TCGGGGCTGC CGCGTATTG CGTCGGGACA CCTTCGCAC
35  201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGCTCGGC ACGGGGGCGA
251 TGCTGCTGCT GTTTACGCG GTAACGCATC TGCCTTTGGC CACCGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCG TTTACACGCA GGCGGTGCTG CTCCTTGTT
40  401 TTGCCGCGT GGTATTGCTG CTTAATCCCT CGTCCGCAG CGGTCAGGAA
451 ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAAGTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTATCCGTT TTGGGCGACG
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
45  651 CATCGGCGTG TCCGCGCTGA TTGCCAACT GTCGATGACG CGCGCTACA
701 AAGTCGCGCA CAAATTCACG GTTGCCCTCG TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCGC ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMLFS
51  TVALGAAAVL RRDTERTPHW KNHLNRSMTG TGAMLLLFYA VTHLPLATGV
101 TLYSTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
55  201 LTGWHTLSFP SAVYLSICIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

```

60  orf135a.pep      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL
      |||
      orf135-1      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL

```

-318-

```

5      orf135a.pep      RRDTFRTPHWKNHLNRSMTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
      orf135-1          RRDXFRTPHWKNHLNRSMTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
      orf135a.pep      RISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
      orf135-1          RISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
10     orf135a.pep      WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
      orf135-1          WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
15     orf135a.pep      VASLSYMTVVFSALSAAFFLAEEFLWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
      orf135-1          VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20 *N.gonorrhoeae*:

```

      orf135.pep      GTGAMLLLFYAVTXLPLATGVTLSTSSIF      30
      orf135ng      STVTLGAAAVLRDTRTPHWKNHLNRSMTGAMLLLFYAVTHLPLTTGVTLSTSSIF      335
25     orf135.pep      LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK      90
      orf135ng      LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLK      395
30     orf135.pep      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM      150
      orf135ng      VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM      455
35     orf135.pep      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVF      201
      orf135ng      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAAF      506

```

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

```

40     1  MPSEKAFRRH  LRTASFQGLH  LHHFHQKV GK  CGIIGFGIHI  FPTLLPAAQG
      51  ILDIQLGLFR  IDFAALAVYR  RTQVDFIHTV  IDGIASDQAF  SEVVQILRRL
      101 NLGHFTDTHL  IAQARRFIAD  FGNIRPMRRG  EAKTFCRCFR  FDGIDGIHGD
      151 FRQCGHINRL  APGKDCRNKG  RDKVFFHTRH  YNQVCLEKTN  CSARKIKFRH
      201 QKQAKTHSTS  LAARFTIRPS  LSQRPFMDTA  KKDILGSGWM  LVAAACFTVM
      251 NVLIKEASAK  FALGSGELVF  WRMLFSTVTL  GAAAVLRDRT  FRTPHWKNHL
      301 NRSMTGAM    LLLFYAVTHL  PLTTGVTLST  TSSIFLAVFS  FLILKERISV
45     351 YTQAVLLLG  AGVLLLNPS  FRSGQEPAL  AGLAGGAMSG  WAYLKVRELS
      401 LAGEPGWRVV  FYLSATGVAM  SSVWATLTGW  HTLSFPSAVY  LSGIGVSALI
      451 AQLSMTRAYK  VGDKFTVASL  SYMTVVFSAL  SAAFFLGEEL  FWQEILGMCI
      501 IISAAF*

```

Further work revealed the following gonococcal sequence <SEQ ID 547>:

```

50     1  ATGGATACCG  CAAAAAAGA  CATTTTAGGA  TCGGGCTGGA  TGCTGCTGGC
      51  GCGCGCCTGC  TTCACCGTTA  TGAACGTATT  GATTAAAGAG  GCATCGGCAA
      101 AATTTGCCCT  CGGCAGCGGC  GAATTGGTCT  TTTGGCGCAT  GCTGTTTTCA
      151 ACCGTTACGC  TCGGTGCTGC  CGCCGTATTG  CGGCGCGACA  CCTTCCGCAC
      201 GCCCATTGG  AAAAACCCT  TAAACCGCAG  TATGGTCGGG  ACGGGGGCGA
55     251 TGCTGCTGCT  GTTTTACGCG  GTAACGCATC  TGCCTTTGAC  AACCAGCGTT
      301 ACCCTGAGTT  ACACCTCGTC  GATTTTttg  GCGGTATTTT  CCTCTCTGAT
      351 TTTGAAAGAA  CGGATTTCCT  TTTACACGCA  GCGGTGCTG  CTCCTTGTTT
      401 TTGCCGGCGT  GGTATTGCTG  CTTAATCCCT  CGTTCGCGAG  CGGTCAGGAA
      451 CCGGCGGCAC  TCGCCGGGCT  GCGGGCGGCG  GCGATGTCCG  GCTGGGCGTA
60     501 TTTGAAAGTG  CGCGAACTGT  CTTTGGCGGG  CGAACCCGGC  TGGCGCGTCG
      551 TGTTTTACCT  TTCCGCAACC  GCGGTGGCGA  TGTCTGcggt  ttgggcgacg
      601 Ctgaccggt  ggCACAcccT  GTCCTTTcca  tgggcagttt  ATCtgtCGGG

```

10

15

20

25

30

35

### Example 66

40

50

55

1 MKRRIAVFVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY  
51 LPGAIEIDSP CGIVEGALLF BHLPAHCLYG KAAVGDAVAH EHPVADVNR

-320-

101 NANAFALFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD  
 151 FDHGKIQQGN NAAAFPPKLA PKIFECFTGA FVGTVYREVC LFYIINDGIA  
 201 HHSAPQVRVY LFAPYCGFLP SASDSLKSS XXSE\*

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

5 1 ATGATGAAGC GGCGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG  
 51 AGTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC  
 101 GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA  
 151 TATCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG  
 201 TCGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCCTGTAT GGTAAAGCCG  
 10 251 CCGTAGGGGA TGCCGTTCGA CACGAACATC CAGTCGCTGA TGTCTCAAC  
 301 CGGAACGCAA ACGCTTTTCG CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT  
 351 CATTGTTCAG CACACCGTAA ATATAAGAC CGTCAAAATA AATATCGTCG  
 401 ATCCACATAT GTTCGCAAAAT TTCGCCGTCT TCGCCGTCTT GGAAGGAGG  
 451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC  
 15 501 AAAAAAGCTC GCGCCAAAAA TATTGAATG TTTACGGGC GCGTTCGTCG  
 551 GCACGGTTTA CCGGTTTCGTC TGCCTGTTCT ACATAATAAA TGACGGAATC  
 601 GCCCATCATT CTGCTCCTCA ACGGTACGG TATCTGTTG CACCTTACTG  
 651 CGGCTTTCTG CCTTCGGCAT CCGATTCCGA TTTGAAAGT TCCAAATATT  
 701 CGGAATAG

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

1 MMKRRIAEV LFPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ  
 51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAVGDAVA HEHPVADVVN  
 101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR  
 151 DFDHGKIQQG NNAAFPPKL APKIFECFTG AFVGTVYREV CLFYIINDGI  
 201 AHHSAPQVRV YLFAPYCGFL PSASDSLKS SKYSE\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

30 orf136.pep 10 20 30 40 50 59  
 MKRRIAEVFLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS  
 orf136a 10 20 30 40 50 60  
 MMKRRIAEVFLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGIAEIDS  
 35 orf136.pep 60 70 80 90 100 110 119  
 PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNNANAFALFDIGQFAXFIVQ  
 orf136a 60 70 80 90 100 110 120  
 PCGIVFGTLLFRHXSTHCLYGKAAVGNAVAHEHPVADVNNANAFALFDIGQFAGFIVQ  
 40 orf136.pep 120 130 140 150 160 170 179  
 HTVNIKTVKINIVDPHMFANFAVFAVLEKRD FDHGKIQQGNNAAFPPKLAPKIFECFTG  
 orf136a 120 130 140 150 160 170 180  
 HAINVKTVKINIVDPHMFANFAXFAVLEKRALTMASKXXMRRSQKSSRQKYLNVLRA  
 45 orf136.pep 180 190 200 210 220 230  
 AFVGTVYREVC LFYIINDGIAHH---SAPQVRVYLFAPYCGFLPSASDSLKSSXXSEX  
 orf136a 180 190 200 210 220 230  
 R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKSSKXSEX

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

55 1 ATGATGAAGC GGCGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG  
 51 GATTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC  
 101 GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA  
 151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG  
 201 TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCCTGTAT GGTAAAGCCG  
 251 CCGTAGGGAA TGCCGTTCGA CACGAACATC CAGTCGCTGA TGTCTCAAC

-321-

5  
301 CGGAACGCAA ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT  
351 CATTTGTTAG CACGCCATAA ATGTAAAGAC CGTCAAATA AATATCGTCG  
401 ATCCACATAT GTTCGCAAAT TTCGCCNTCT TCGCCGTCTT GGAAAAAAGG  
451 GCTTTGACCA TGGCAAAATC TAAGNGNNA NNGATGCGGC GGC GTTCCCA  
501 AAAAAGCTCG CGCCAAAAAT ATTGAATGT TTTGCGGGCG CGTTCGCCGG  
551 CACGGTTTAC CGGTTTGTCT GCCTGTTCTA CATAATAAAT GACGGAATCG  
601 CCCATCATAT CTGCTCCTCA ACGGTACGG TATCTGTTT CACCTTACTG  
651 CGGCTTCTG CCTTCGGCAT CCGATTCGGA TTTGAAAAGT TCCAAATATT  
701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

15  
1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHQQ  
51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAGVNAVA HEHPVADVNN  
101 RNANAFALFD IGQFAGFIVQ HAINVKT VKI NIVDPHMFAN FAXFAVLEKR  
151 ALTMASKXX XMRRRSQKSS RQKYNLNLRA RSPARFTGLS ACST\*\*MTES  
201 PIISAPQVRV YLFAPYCGFL PSASDSDLKS SKYSE\*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

20  
orfl36a.pep 10 20 30 40 50 60  
MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGLAEIDS  
orfl36-1 10 20 30 40 50 60  
MMKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLAEIDS  
25  
orfl36a.pep 70 80 90 100 110 120  
PCGIVFGTLLFRHXSTHCLY GKAAGVNAVAHEHPVADVNNRNANAFALFDIGQFAGFIVQ  
orfl36-1 70 80 90 100 110 120  
PCGIVFGALLFRHLPAHCLY GKAAGVDAVAHEHPVADVNNRNANAFALFDIGQFAGFIVQ  
30  
orfl36a.pep 130 140 150 160 170 180  
HAINVKT VKI NIVDPHMFANFAXFAVLEKRALTMASKXXXMRRRSQKSSRQKYNLNLRA  
orfl36-1 130 140 150 160 170 180  
HTVNIKT VKI NIVDPHMFANFAVFAVLEKRD FDHGKI QGGNNAAFPKKLAPKIFECFTG  
35  
orfl36a.pep 190 200 210 220 230  
R---SPARFTGLSACSTXXMTESPIISAPQVRVRYLFAPYCGFLPSASDSDLKSSKYSEX  
orfl36-1 190 200 210 220 230  
AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVRYLFAPYCGFLPSASDSDLKSSKYSEX

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from *N.gonorrhoeae*:

45  
orfl36.pep MKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLAEIDS 59  
orfl36ng MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGLAEIDS 60  
50  
orfl36.pep PCGIVFGALLFRHLPAHCLY GKAAGVDAVAHEHPVADVNNRNANAFALFDIGQFAXFIVQ 119  
orfl36ng PGGIVFGTLLFRHLSAHCLY GKAAGVDAVAHEHPVADVNNRNANAFALFDIGQSAGFIVQ 120  
55  
orfl36.pep HTVNIKT VKI NIVDPHMFANFAVFAVLEKRD FDHGKI QGGNNAAFPKKLAPKIFECFTG 179  
orfl36ng HTVNIKT VKI NIVDPHMFANFAVFAVLEKRD FDHGKI QGGNNAAFPKKLAPKVFECFTG 180  
orfl36.pep AFVGTVYRFVCLFYIINDGIAHHSAPQVRVRYLFAPYCGFLPSASDSDLKSSXXSE 234  
orfl36ng AFAGTVYRFVCLFYIINDGIAHHTAPQVRVRYLFAPYRGFLPPASDSDLKSSKYSE 235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60  
1 ATGATGAAGC GCGTATAGC CGTCTCGTC CTGCTCATGC AGAAAATCCG  
51 GATTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC

-322-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

```

101 GGATGCTCTT CCAAATTTTC GGGATGTTCT TTTTCTTCAT ACACCGGCAA
151 TACCTGCCCC GGATCGCCGA AATCGATTCC CCAGGCGGTA TCGTGTTCGG
201 TACGCTCCTC TTCCGTCATC TGTCCGCGCA TTGCCTGTAC GGTAAAGCCG
251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGCCAAC
301 CGGAACGCAA ACGCTTTTCG CTTGTTTCGAC ATTGGTCAGT CCGCCGGGTT
351 CATTGTTTCA CACACCGTAA ATATAAAGAC CGTCAAAATA AATATCGTCG
401 ATCCACATAT GTTCGCAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAAGG
451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC
501 AAAAAAGCTC GCGCCAAAAG TATTTGAATG TTTTACGGGC GCGTTCGCGC
551 GCACGGTTTA CCGGTTTCGT TGCCTGTTCT ACATAATAAA TGACGGAATC
601 GCCCATCATA CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACCG
651 CCGTTTTCTA CCTCCGGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT
701 CGGAATAG

```

This encodes a protein having amino acid sequence <SEQ ID 556>:

15  
20  
25  
30  
35

```

1 MMKRRIAEFV LLMQKIRILG QLLPKIVNTV PAHRMLFQIF GMFFFFIHRQ
51 YLPGLAEIDS PGGIVFGTLL FRHLSAHCLY GKAAVGDAVA HEHPVADVAN
101 RNANAFALFD IGQSAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
151 DFDHGKIQGG NNAAAFPPKL APKVFECFTG AFAGTVYREV CLFYIINDGI
201 AHHTAPQVRV YLFAPYRGFL PPASDSLKS SKYSE*

```

20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

25  
30  
35

```

orf136ng      MMKRRIAEFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGLAEIDS
                |||||:| ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf136-1      MMKRRIAEFVLFPPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLAEIDS

orf136ng      PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVANRNANAFALFDIGQSAGFIVQ
                | ||||:||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
orf136-1      PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNVNANAFALFDIGQFAGFIVQ

orf136ng      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPPKLAPKVFECFTG
                |||||:||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
orf136-1      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPPKLAPKIFECFTG

orf136ng      AFAGTVYREVCLFYIINDGIAHHTAPQVRVYLFAPYRGFLPPASDSLKSSEK
                ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf136-1      AFVGTVYREVCLFYIINDGIAHHSAPQVRVYLFAPYCGFLPSASDSLKSSEK

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 67

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

45  
50

```

1 ATGGAATA TGTAACGTT TTCAAAATC AGACCGCTTT TGGCAATCGC
51 CGCCGCCGCG TTGCTTGCCG CC.TGCGGAC GGCGGAAAT AATGCTGTCC
101 GCAAGCCGGT GCAAACCGCC AAACCGCGCG CAGTGGTCGG TTTGGCACTC
151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
201 GAAAGAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACC TCCGCAGGTT
251 CGATTGTCGG CAACCTTTT GCATCGGGTA TGTCGCCGA CCGCCTCGAA
301 TTGAAGCCG AATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
351 CACCAATGGG TTTATCAAAG GCGCAAAGCT GCAAAATAC ATCAACCGAA
401 AACTCCGCG CATGCAGATT CAGCAGTTT CCATCAAATT TGCCGCC..

```

50 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

```

1 MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL
51 GGGASKGFAH VGIKVLKEN GIPVKVTGT SAGSIVGNLF ASGMSPDRL
101 LEAEILGKTD LVDLTSTNG FIRKAKLQNY INRKLGMQI QQFPIKFAA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 559>:



-323-

1 ATGGAATA TGGTAACGTT TTCAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGTCC  
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 5 201 GAAAGAAAC GGTATTCTCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
 251 CGATTGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 10 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCCGCGTT  
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AGGGGAATGC  
 501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG  
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG  
 601 CCCGTCAGTG CCGCCCGGCG GCAGGGGGCG AATTTCTGTA TTGCCCTCGA  
 15 651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC  
 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTGTGA AAATGAGTTG  
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT  
 801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTGGGT GAGGAGCAG  
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL  
 51 GGGASKGFAH VGIIVKLVEN GIPVKVVTGT SAGSIVGSLF ASGMS PDRLE  
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV  
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV  
 201 PVSAAARRQGA NFVIAVDISA RPKNISQGF FSYLDQTLNV MSVSALQNEL  
 251 QQADVVIKPQ VLDLGA VGGF DQKKRAIRLG EEAAARAALPE IKRKLAAARY  
 301 \*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf137.pep		MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTA	KPAAVVGLALGGGASKGFAH				
35	orf137a	MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTA	KPAAVVGLALGGGASKGFAH				
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf137.pep	VGIIVKLVKENGIPVKVVTGTSAGSIVGNLFASGMS PDRLELEAEILGKTDLVDLTLSTNG					
	orf137a	VGIIVKLVKENGIPVKVVTGTSAGSIVGSLFASGMS PDRLELEAEILGKTDLVDLTLSTSG					
		70	80	90	100	110	120
		130	140	149			
45	orf137.pep	FIKGAKLQNYINRKLGRMQIQFPIKFAA					
	orf137a	FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV					
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 1 ATGGAATA TGGTAACGTT TTCAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGTCC  
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 55 201 GAAAGAAAC GGTATTCTCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
 251 CGATAGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG TAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCCGCGTT  
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGAATGC  
 60 501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG  
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG

-324-

5  
601 CCGGTCAAGT CCGCCCGCG GCANGNNNG NATNTCGTGA TTGCCGTCGA  
651 TATTTCGCC CGTCCGAGCA AAAACATCAG CCAAGGCTTC TTCTCTTATC  
701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTGCA AAATGAGTTG  
751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT  
801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTGGGT GAGGAGGCAG  
851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10  
1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL  
51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE  
101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRRI QQFPIKFAAV  
151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV  
201 PVSAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL  
251 GQADVVIKQP VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY  
15 301 \*

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20  
orfl37a.pep MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH  
orfl37-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH  
orfl37a.pep VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG  
orfl37-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG  
25  
orfl37a.pep FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
orfl37-1 FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
30  
orfl37a.pep FQPVIIGRHTYVDGGLSQPVVSAARRXXXXXVIAVDISARPSKNISQGFSSYLDQTLNV  
orfl37-1 FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGFSSYLDQTLNV  
orfl37a.pep MSVSALQNELGQADVVIKQPQLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY  
35 orfl37-1 MSVSALQNELGQADVVIKQPQLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

40  
orfl37.pep MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH 60  
orfl37ng MENMVTFSKIRSFLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH 60  
45  
orfl37.pep VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTLSTNG 120  
orfl37ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMSPDRLELEAEILGKTDLVDLTLSTSG 120  
orfl37.pep FIKGAKLQNYINRKLGRMQIQFPIKFAA 149  
50 orfl37ng FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

55  
1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGATCATTTT TGGCAATCGC  
51 CGCCGCCGCG TTGCTTGCCG CCTGCGGTAC GGCGGGAAAC AATGCCGCC  
101 GCAAGCCGGT GCAAACCGCC AAACCGCCG CAGTGGTCG TTTGGCACTC  
151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT ATAGGAATTG TTAAGGTTTT  
201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
251 CGATAGTCGG CAGCCTTTTG GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
301 TTGGAAGCCG AGATTTTAGG TAAAACCGAT TTAGTCGATT TAACCTGTCT  
351 CACCAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
60 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT  
451 GCCACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGAATGC

-325-

501 CGGGCAGGCG GTTCGTGCTT CCGCCGCCAT TCCCAATGTG TTCCAGCCAG  
 551 TCATCATCGG CAGGCACAAA TATGTTGACG GCGGTCTGTC GCAGCCCGTG  
 601 CCCGTCAGTG CCGCTCGGCG GCAGGGGGCG AATTTCTGTA TTGCCGTCGA  
 651 TATTTCCGCA CGTCCGAGCA AAAATGTCGG TCAAGGTTTC TTCTCTTATC  
 701 TCGATCAGAC GCTGAACGTG ATGAGCGTTT CCGTGTGCA AAACGAGTTG  
 751 gggcAGGCGG ATGTGGTTAT CAAACCGCag gtTTTGGATT TGGGTGCAGT  
 801 CGGCGGATTC GATCAGAAAA AGCGCGCCAT CCGGTTGGGC GAGGAGGCAG  
 851 CACGTGCCGC ATTGCCGTAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL  
 51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSL LSGMSPDRLE  
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV  
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV  
 15 PVSAARRQGA NFVIAVDISA RPSKNVGQGF FSYLDQTLNV MSVSVLQNEL  
 251 GQADVVIKPQ VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY  
 301 \*

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVTFSKIRSFLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH  
 orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH  
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMSPDRLELEAEILGKTDLVDLTLSTSG  
 25 orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG  
 orf137ng FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
 orf137-1 FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
 30 orf137ng FQPVIIGRHKYVDGGLSQPVVSAARRQGANFVIAVDISARPSKNVGQGFSSYLDQTLNV  
 orf137-1 FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGFSSYLDQTLNV  
 35 orf137ng MSVSVLQNELGQADVVIKPQVLDLGA VGGFDQKKRAIRLGEAARAALPEIKRKLAAARY  
 orf137 MSVSALQNELGQADVVIKPQVLDLGA VGGFDQKKRAIRLGEAARAALPEIKRKLAAARY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site  
 (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and  
 40 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for  
 raising antibodies.

### Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGcTG CCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCmAT ATGCGGCAGG CGGGTTTGAA  
 201 CCCCAGCCCC AAAACGCTCA AAGCCGTTT TGCGGAAACG GCAAAAAGGCG  
 251 GTTTGGAAct TGCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA  
 50 301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA  
 351 ACACGAAGGG CTGCTATTC..

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL LPLSCLHTLGN RLGHLEAFYLL  
 55 51 KEDRARIVAX MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDIET  
 101 MFKAVHGWEH VQQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCTGCTG  CCGCTTTCCT
   101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
    15  151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATCGGCGAGG  CGGGTTTGAA
      201  CCCCACCCC  AAAACGGTCA  AAGCCGTTT  TCGGAAACG  GCAAAAGGCG
      251  GTTTGGAAC  TGCCCCCGCG  TTTTTCAGAA  AACCGGAAGA  CATAGAAACA
      301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGGACAA
      351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTTGG
     10  401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
      451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
      501  TCGCGGCAAA  GGA AAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAACAAA
      551  TCATCAAAGC  CCTGCGTTTC  GGCGAAGCAA  CCATCGTCCT  GCCGACCAC
      601  GTCCCTCCC  CTCAAGAAGG  CGGGGAAGGC  GTATGGGTGG  ATTTCTTCGG
     15  651  CAAACCTGCC  TATACCATGA  CGCTGGCGGC  AAAATTGGCA  CACGTCAAAG
      701  GCGTGAAAC  CTGTTTTTTC  TGCTGCGAAC  GCCTGCCTGG  CGGACAAGGT
      751  TTCGATTGTC  ACATCCGCC  CGTCCAAGGG  GAATTGAACG  GCGACAAAGC
      801  CCATGATGCC  GCCGTGTTCA  ACCGCAATGC  CGAATATTGG  ATACGCCGTT
      851  TTCCGACGCA  GTATCTGTTT  ATGTACAACC  GCTACAAAAT  GCCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

      1  MFRLQFRLFP  PLRTAMHILL  TALLKCLSL  PLSCLHTLGN  RLGHlafyLL
    51  KEDRARIVAN  MRQAGLNPD  KTVKAVFAET  AKGGLLELAP  FFRKPEDLET
   101  MFKAVHGWHE  VQALDKHEG  LLFITPHIGS  YDLGGYISQ  QLPFPLTAMY
    25  151  KPPKIKAIK  IMQAGRVRGK  GKTAPTSIQG  VKQIKALRS  GEATIVLPDH
      201  VSPQEGGEG  VWVDFFGKPA  YTMTLAAKLA  HVKGVKTLFF  CCELRPGGQG
      251  FDLHIRPVQG  ELNGDKAHDA  AVENRNEYW  IRRFPTQYLF  MYNRYKMP*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N. meningitidis*:

```

                                     10      20      30      40      50      60
orfl38.pep  MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHlafyLLKEDRARIVAX
orfl38a     MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHlafyLLKEDRARIVAN
                                     10      20      30      40      50      60
                                     70      80      90     100     110     120
orfl38.pep  MRQAGLNPDPKTVKAVFAETAKGGLLELAPAFFRKPEDLETMFKAVHGWHEHVQALDKHEG
orfl38a     MRQAGLNPDPKTVKAVFAETAKGGLLELAPAFFRKPEDLETMFKAVHGWHEHVQALDKHEG
                                     70      80      90     100     110     120

orfl38.pep  LLF
orfl38a     LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG
                                     130     140     150     160     170     180

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    50  51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCTGCTG  CCGCTTTCCT
      101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
      151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGTCAGG  CAGGCATGAA
      201  TCCCACCCC  AAAACGGTCA  AAGCCGTTT  TCGGAAACG  GCAAAAGGCG
      251  GTTTGGAAC  TGCCCCCGCG  TTTTTCAGAA  AACCGGAAGA  CATAGAAACA
     55  301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGGACAA
      351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTTGG
      401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
      451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
      501  TCGCGGCAAA  GGA AAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAACAAA
     60  551  TCATCAAAGC  CCTGCGTTTC  GGCGAAGCAA  CCATCGTCCT  GCCGACCAC

```

-327-

5  
601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG  
651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG  
701 GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT  
751 TTCGATTTCG ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC  
801 CCATGATGCC GCCGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT  
851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

10  
1 MFRLLQFRLFP PLRTAMHILL TALLKCLSL PLSCLHTLGN RLGHLAFYLL  
51 KEDRARIVAN MRQAGLNPD KTVKAVFAET AKGGLLELAPA FFRKPEDIET  
101 MFKAVHGW EHVQQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY  
151 KPPKIKAIK IMQAGRVRGK GKTAPTSIQG VKQIIKALRS GEATIVLPDH  
201 VPSPQEGGEG VWVDFFGKPA YTMTLAAKLA HVKGVKTLFF CCERLPGGQG  
251 FDLHIRPVQG ELNGDKAHD AAVFNRAEYW IRRFPTQYLF MYNRYKMP\*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15 orf138a.pep MFRLLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN  
orf138-1 MFRLLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN  
20 orf138a.pep MRQAGMNPDPKTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHEG  
orf138-1 MRQAGLNPD KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHEG  
orf138a.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG  
25 orf138-1 LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG  
orf138a.pep VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF  
orf138-1 VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF  
30 orf138a.pep CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRAEYWIRRFPTQYLFMYNRYKMP  
orf138-1 CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRAEYWIRRFPTQYLFMYNRYKMP

### 35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from *N.gonorrhoeae*:

40 orf138.pep MFRLLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX 60  
orf138ng MFRLLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN 60  
orf138.pep MRQAGLNPD KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHEG 120  
orf138ng MRQAGLNPD KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHEG 120  
45 orf138.pep LLF 123  
orf138ng LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG TCGCTTTCCT  
101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA  
201 CCCCAGACAG CAGACGCTCA AAGCCGTTTT TGCGGAAACG GCAAATGCG  
55 251 GTTTGGAAGT TGCCCCCGCG TTTTCAAAA AACCAGGAAGA CATCGAAACA  
301 ATGTTCAAAG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGACAA  
351 GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGG  
401 GCGGACGCTA CATCAGCCAG CAGCTCCGT TCCACGTGAC CGCCATGTAC  
451 AAGCCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT  
60 501 GCGCGGCAAA GGCAAAACcg cgccaccgg catACAAGGG GTCAAACAAA  
551 tcatcaAGGC CCTGCGCGCG GCGGAGGCAA CCAAtcATCCT GCCCGACCAC

-328-

5  
 601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA  
 651 ACCTGCATAc acCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG  
 701 TGAAAACCCCT GTTTTTCTGC TGCGAACGCC TGCCCGACGG ACAAGGCTTC  
 751 GTGTTGCACA TCCGCCCCGT CCAAGGGGAA TTGAACGGCA ACAAAGCCCA  
 801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTC  
 851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

10  
 1 MFRLQFRLFP PLRTAMHILL TALLKCLSL SLSCSLHTLGN RLGHlafyLL  
 51 KEDRARIVAN MRQAGLNPDt QTVKAVFAET AKCGLELAPA FFKKPEDiet  
 101 MFKAVHGWEH VQqALDKGEG LLFITPHIGS YDLGGRYISQ QLpFHLTAMY  
 151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQg VKQIIKALRA GEATIILPDH  
 201 VPSPQEGGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDQGGF  
 251 VLHIRPVQGE LNGNKAHDAA VENRNTEYWI RRFPTQYLFM YNRYKTP\*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15 orf138-1.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSLCLHTLGNRLGHlafyLLKEDRARIVAN  
 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSLSLSCSLHTLGNRLGHlafyLLKEDRARIVAN  
 20 orf138-1.pep MRQAGLNPDPKTVKAVFAETAKGGLLELAPAFFKKPEDietMFKAVHGWEHVQqALDKHEG  
 orf138ng MRQAGLNPDtQTVKAVFAETAKCGLELAPAFFKKPEDietMFKAVHGWEHVQqALDKGEG  
 25 orf138-1.pep LLFITPHIGSYDLGGRYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTISIQg  
 orf138ng LLFITPHIGSYDLGGRYISQQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQg  
 30 orf138-1.pep VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFGKPAYTMTLAAKLAHVKGvKTLFF  
 orf138ng VKQIIKALRAGEATIILPDHVPSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGvKTLFF  
 35 orf138-1.pep CCERLPGGQGFDLHIRPVQGELNGDKAHDAAVFNRAEYWIRRFPTQYLFMYNRYKMP  
 orf138ng CCERLPDQGGFVLHIRPVQGELNGNKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35 gnl|PID|e334283 (Y14568) htrB [*Pseudomonas fluorescens*] Length = 253  
 Score = 80.8 bits (196), Expect = 9e-15  
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)  
 40 Query: 101 MFKAVHGWEHVQqALDKGEGLLFITPHIGSYD-LGGRYISQQLPFHLTAMYKPPKIKAIK 159  
 + + V G E ++AL G+G++ IT H+G+++ L Y SQ P Y+PPK+KA+D  
 Sbjct: 94 LVREVEGLEVLKEALASGKGvGITSHLGNWEVLNHFYCSQCKPI---IFYPKPKLKAVD 150  
 Query: 160 KIMQAGRVRGKGKTAPTGTGIQGVKQIIKALRAGEATIILPDHVPSPQEGGvGVWADFFGKPA 219  
 ++++ RV+ K A + +G+ +IK +R G I D P P E G++ FF A  
 45 Sbjct: 151 ELLRKQRVQLGNKVAASKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVPFFATQA 208  
 Query: 220 YTMTLAAKLAHVKGvKTLFFCCERLPDQGGF 250  
 T + +F RLPDG G+  
 Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the  
 gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.  
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A  
 55 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein  
 was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5      1  ..GCGTGGTCGG CCGGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
      51  GCATGCGGTG TGAATACTT TGCCTTCTC GCGGCGGCG GTGTATGCGG
     101  CAGCGGTTT GGTGTGGTG TATGCGGCGC GCGGCGGCG GTGCGCGTGG
     151  ATGCGCGGGC TGATGTTTTA GCCGTTTATG GTGTCGCCGG TTTGTGTTTC
     201  GCGGGGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTCG TTGCCGTTGC
    10  251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
     301  TTATCAGCCT GGGATGCACT GCCGCCGAT TACGGCAGGG CCGCGGCGGG
     351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCAGC TTCCCCCTCT
     401  TGAAACCGGC GTTGCGGCGC GGTCTGACTT TGGCGGCGGC AACCTGCGTG
     451  GGCGAATTG CGCGACATT GTTCTGTGCG CGTCCGGAAT GGCAGACGCT
    15  501  GACGACTTGG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG
     551  CGCGGGCGAT GGTGCTG..

```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

      1  ..AWSAGESWRV LMESETWHAV WNTLRFSAAS VYAAAVLGVV YAAPARRSAW
     51  MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
    20  101  LSAWDALPPD YGRAAGLGA NGFQTACRIT FPLLKPALRR GLTLAAATCV
     151  GEFAATFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..

```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

      1  ATGGATGGAC GCGCTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
     51  GGCTTTTFTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
    25  101  ATGACGGTTT GCGTGGGCGC GCGTGCTGT CGGATGCCTA TATGCTCAAA
     151  CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
     201  GCCTTTGGGC GTGCCTGTGCG CGTGGGTGCT GCGCGGCTG GCGTTTCCGG
     251  GCGGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTGTG GATGCCACAG
     301  TTGGTGGCGG GCGTGGGCGT GCTGCGCCGT TTCGGGGCGG ACGGGCTGTT
    30  351  GTGGCGCGGC AGGCAGGATA CGCCGTATCT GTTGTGTGAC GGCAATGTGT
     401  TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTGTGCAAA
     451  GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CCGGGGCGTG
     501  GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
     551  GCGGCGGTG CTTGTCTTTT CTGTATTGTT TTTCCGGGTT CCGGCTGGCG
    35  601  CTGCTGCTGG GCGGCGAGCG TTATGCCACG GTCGAAGTGG AAATTTACCA
     651  GTTGGTCATG TTCGAACTCG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC
     701  TGGTGTGGG GGTAAACGGC GCGGCAGGGT TGCTGTATGC GTGGTTCGCG
     751  AGGCGCGCGG TTTCGGATAA GGCGGTTTCC CCTGTGATGC CGTCGCGGCC
     801  GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTTGCGGCG GCGGTGTTGT
    40  851  CTGTGTGCTG CCTGTTTCCT TTGTTGGCAA TTGTTGTGAA AGCGTGGTCG
     901  GCCGCGCAAT CGTGGCGTGT GTTAATGGAA AGTGAACGT GGCAGGCGGT
     951  GTGGAATACT TTGCGCTTCT CCGCGGCGGC GGTGTATGCG GCGGCGGTTT
    1001  TGGGTGTGGT GTATGCGGCG GCGGCGCGCG GGTGCGCGTG GATGCGCGGG
    1051  CTGATGTTTT TGCCGTTTAT GGTGTCGCGG GTTGTGTTTT CCGCGGGCGT
    45  1101  GCTGCTGCTT TATCCGCACT GGACGGCTTC GTTGCCGTTG CTGCTGGCGA
     1151  TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC
     1201  TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCGG GTTTGGGTGC
     1251  AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG
     1301  CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
    50  1351  GCGGCGACAT TGTTTCTGTC GCGTCCGAA TGGCAGACGC TGACGACTTT
     1401  GATTTATGCC TATTGGGAC GCGCGGGTGA GGATAATTAC GCGCGGGCGA
     1451  TGGTGCTGAC ATTGCTGTTG GCGGCGTTCG CGCTGGGTAT TTTCTGCTG
     1501  TTGGACGGCG GCGAAGGCGG AAAACAGACG GAAACGTTAT AA

```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55      1  MDGRWVVWG AFALLPSAFL AVMVVAPLWA VAAYDGLAWR AVLSDAYMLK
     51  RLAWTVFQAA ATCVLVPLG VPVAWVLARL AFPGRALVLR LLMLPFVMP
    101  LVAGVGVLAL FGADLLWRG RQDTPYLLLY GNVFENLPVL VRAAYQGFVQ
     151  VPAARLQ TAR TLGAGAWRRF WDIEMFVLRP WLAGGVCLVF LYCFSGFGLA

```

-330-

201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVLGVTA AAGLLYAWFG  
 251 RRAVSDKAVS PVMPSPPQSV GEYVLLAFAA AVLSVCCLFP LLAIVVKAWS  
 301 AGESWRVIME SETWQAVWNT LRFSAAAVYA AAVLGVVYAA AARRSAWMRG  
 351 LMFLPFMVSP VCVSAGVLLL YPQWTASLPL LLAMYALLAY PFVAKDVLSA  
 401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF  
 451 AATLFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL AAFALGIFLL  
 501 LDGGEKGKQT ETL\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N. meningitidis*:

					10	20	30
	orf139.pep				AWSAGESWRVLMES	ETWHAVWNTLRFS	AAA
15	orf139a	QSVGEYVLLAF	AAVXSVCC	FLXLLAIVVK	AWSSAGESWRVLMES	ETWQAVWNTXRF	SA
		270	280	290	300	310	320
	orf139.pep		40	50	60	70	80
			VYAAAVLG	VYAAPARR	SAWMRGLMFX	PFMVSPVCV	SAGVLLLYPQWTASL
20	orf139a		VYAAAVLG	VYAAAARR	SAWMRGLMFL	PFMVSPVCV	SAGVLLXPQWTASL
			330	340	350	360	370
	orf139.pep		100	110	120	130	140
			LAYPFVAKDVLSA	WDALPPDYGRA	AAGLGANGFQTACRIT	FPLLKPALRRGLT	LAAATCV
	orf139a		LAYPFVAKDVLSA	WDALPPDYGRA	AAGLGANGFQTACRIT	FPLLKPALRRGLT	LAAATCV
			390	400	410	420	430
30	orf139.pep		160	170	180	189	
			GEFAATLFLSRPE	WQTLTTLIYAYL	GRAGEDNYARAMV	L	
	orf139a		GEFAATLFXSRX	EWQTLTTLIYAYX	GRAGXDNYARAMV	LTLAFAALGXFL	LDGGE
			450	460	470	480	490
							500

35 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

	1	ATGGATGGAC	GGCGTTGGGC	GGTATGGGGT	GCTTTTGCCC	TGCTGCCTTC
	51	GGCTTTTGTG	GCGGCAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGCGT
	101	ATGACGGTTT	GGCGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAA
	151	CGTTTGGCGT	GGACGGTATT	TCAGGCAGCG	GCAACCTGTG	TGCTGGTGCT
40	201	GCCTTTGGGC	GTGCTGTGCG	CGTGGGTGCT	GGCGGGGCTG	GCGTTTCCGG
	251	GGCGGGGCTT	GGTGTGCGCG	CTGCTGATGC	TGCCTTTTGT	GATGCCACG
	301	TTGGTGGCGG	GCGTGGGCGT	GCTGGCTCTG	TTGGGGGCGG	ACGGCCTGTN
	351	GTGGCGCGGC	TGGCAGGATA	CGCCGTATCT	GTTGTGTGAC	GGCAATGTGT
	401	TTTTTNACCT	TCCTGTGTTG	GTCAGGGCGG	CATATCAGGG	GTTTGTGCAA
45	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACNG	ACATTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCGT	TTTGGGCCCC	TGGCTTGCCG
	551	GCGGCGGTG	CCTTGCTTTC	CTGTATTGTT	TTTGGGGGTT	CGGGCTGGCA
	601	TTGCTGTGCG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGC	AAATTTACCA
	651	GTTGGTCATG	TTCGAAGTGC	ATATGGCGGT	TGCTTCGGTG	CTNGTGTGGC
50	701	TGGTGTNGGG	GGTAACNGCG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGGCGCGCGG	TTTGGGATAA	GGCNGTTTCC	CCTGTGATGC	CGTCGCGGCC
	801	GCAGTGCGTC	GGGGAATATG	TGCTNCTGGC	GTTTGGGCGG	GCGGTGTNGT
	851	CTGTGTGCTG	CCTGTTTCNT	TTGTTGGCAA	TTGTTGTGAA	AGCGTGTGCG
	901	GCGGCGCAAT	CGTGGCGTGT	GTTAATGGAA	AGTGAACGCT	GGCAGGCGGT
55	951	GTGGAATACT	NTGCGCTTCT	CGGCGGCGGC	GGTGTATGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGGCGG	GCTCGGCGTG	GATGCGGCGG
	1051	CTGATGTTTT	TGCCGTTTAT	GGTGTGCGCG	GTTTGTGTTT	CGGCGGCGGT
	1101	GCTGCTGCTT	NATCCGCACT	GGACGCGCTT	GTTGCCGCTG	CTGCTGGCGA
	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCAGCC
60	1201	TGNGATGCAC	TGCCGCGCGA	TTACGGCAGG	GCGGCGGCGG	GTTTGGGTGC
	1251	AAACGGGCTT	CAGACGGCAT	GCCGCATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	CGGTCTGACT	TTGGCGGCGG	CAACCTGCGT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTCTNTGTC	GCGTCNCGAG	TGGCAGACGC	TGACGACTTT



5	1	MDGRRWAVWG	AFALLPSAFL	AAMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	<u>ATCVLVLP</u> LG	VPVAVLRL	AFPGRALVLR	LLMLPFVMPT
	101	LVAGVGLAL	FGADGLXWRG	WQDTPYLLLY	GNVFFXLPVL	VRAAYQGFVL
	151	VPAARLQTAX	FLGAGAWRRF	WDIEMPVLRP	LVAGGVCLVF	LYCFCSGFLA
	201	LLLGGSRYAT	VEVEIYQLVM	FELDMAVASV	<u>LVWLXGVTA</u>	AAGLLYAWFG
10	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFAA	AVXSVCCFLX	LLAIVVKAWS
	301	AGESWRVLME	SETVQAVWNT	XRFSAAAVYA	<u>AAVLGVVYAA</u>	AARRSAWMRG
	351	LMFLPFMVSP	CVVSAGVLLL	XPQWTASLPL	LLAMYLALLAY	PFVAKDVLSA
	401	XDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFXSRXE	<u>WQTLTTLIYA</u>	YXGRAGXDNY	ARAMVLTLLL	<u>AAAFALGXFL</u>
15	501	LDGGEKKRT	ETL*			

	orf139a.pep	MDGRRWAVWGAFALLPSAFLAAMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA      :     :
20	orf139-1	MDGRRWVVGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
	orf139a.pep	ATCVLVLPGLGVPVAWVLARLAFPPGRALVRLRLMLPFVMPPTLVAGVGV LALFGADGLXWRG 
	orf139-1	ATCVLVLPGLGVPVAWVLARLAFPPGRALVRLRLMLPFVMPPTLVAGVGV LALFGADGLLWRG
25	orf139a.pep	WQDTPYLLLYGNVFFXLPVLVRAAYQGQVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP 
	orf139-1	RQDTPYLLLYGNVFFNLPVLVRAAYQGQVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
30	orf139a.pep	WLAGGVCLVFLYCFSGFGLALLGGSR YATVEVEIYQLVMFELDMAVASVLVWLXGVTA 
	orf139-1	WLAGGVCLVFLYCFSGFGLALLGGSR YATVEVEIYQLVMFELDMAVASVLVWLVLGVT A
	orf139a.pep	AAGLLYAWFGRRAVSDKAVSPVMPSP PQSVGEYVLLAFAAAVXSVCCFLXLLAI VVKAWS 
35	orf139-1	AAGLLYAWFGRRAVSDKAVSPVMPSP PQSVGEYVLLAFAAAVLSVCCFLP LLAIVVKAWS
	orf139a.pep	AGESWRVLMSESETWQAVWNTXRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP 
40	orf139-1	AGESWRVLMSESETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
	orf139a.pep	VCVSAGVLLLPQWTASLP LLLAMYALLAYPFVAKDVL SAXDALPPDYGRAAAGLGANGF 
	orf139-1	VCVSAGVLLLPQWTASLP LLLAMYALLAYPFVAKDVL SAWDALPPDYGRAAAGLGANGF
45	orf139a.pep	QTACRITFPL LKPALRRGLTLAAATCVGEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNY 
	orf139-1	QTACRITFPL LKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
50	orf139a.pep	ARAMVLTLLLA AFALGXFLLLDGEGGKRTETLX 
	orf139-1	ARAMVLTLLLA AFALGIFLLLDGEGGKOTETLX

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

	orf139.pep	AWSAGESWRVLMSESETWHAVWNTRLRFSAAA	30
		:	
	orf139ng	QSVGEYVLLAFSVAVLVSVCCFLPLSAIVVKAWSAGESRRVLMSESETWQAVWNTRLRFSAAA	327
60	orf139.pep	VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLAMYAL	90
		:       :	
	orf139ng	VFAAAVLGVVYAAARRLVWMRGLVFLPFMVSPVCVSAGVLLLYPGWTASLPLLAMYAL	387

orf139.pep	LAYPFVAKDVL <del>SAWDALPPDY</del> GRAAAGLGANGFQTACRITTFLLKPALRRGLT <del>LAAATCV</del>	150
orf139ng	LAYPFVAKDVL <del>SAWDALPPDY</del> GRAAAGLGANGFQTACRITTFLLKPALRRGLT <del>LAAATCV</del>	447
5 orf139.pep	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL	189
orf139ng	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL <del>TLLLSAFAVCIFLLLDNGEGG</del>	507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

10	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLPPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPPT
	101	LVAGVGV <del>LAL</del>	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPLV	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRPF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSR <del>YAT</del>	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
15	251	RRAVSDKAVS	PVMPSPPPQSV	GEYVLLAFSV	AVLSVCCLFP	LSAIVVKAWS
	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVL <del>SA</del>
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
20	501	LDN <del>EGG</del> GKRT	ETL*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

	1	ATGGATGGAC	GGTGTGGGCG	GGTACGGGGT	GCTTTTTCCT	TGCTGCCTTC
	51	GGCTTTTTCG	GCGGTAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGGCGT
	101	ATGACGGTTT	GGCGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
25	151	CGTTTGGCGT	GGACGGTGT	TCAGGCGGCG	GCAACCTGTG	TGCTGGTGCT
	201	GCCTTTGGGC	GTGCCTGTCG	CGTGGGTGCT	GGCGCGGCTG	GCGTTCCCGG
	251	GGCGGGCTTT	GGTGCTGCGC	CTGCTGATGC	TGCCGTTTGT	GATGCCACAC
	301	CTGGTGGCGG	GCGTGGGCGT	GCTGGCTCTG	TTCCGGGCGG	ACGGGCTGTT
	351	GTGGCGCGCG	CGGCAGGATA	CGCCGTATCT	GTTGTTGTAC	GGCAATGTGT
30	401	TTTTCAACCT	CCCCGTGTTG	GTCAGGGCGG	CGTATCAGGG	GTTTGCTCAA
	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACGG	ACGTTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCGT	TTTGCGCCCG	TGGCTTGCCG
	551	GCGGCGGTGG	CCTGTCTTTC	CTGTATTGTT	TTTCCGGGTT	CGGGCTGGCA
	601	TTGCTGTTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
35	651	GTTGGTTATG	TTCGAACTCG	ATATGGCGGG	GGCTTCGGCG	CTGGTGTGGC
	701	TGGTGTGGG	GGTAACGGCG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGGCGCGCGG	TTTCGGATAA	GCGGTTTTC	CCCGTGATGC	CGTCGCCCGC
	801	GCAATCGGTG	GGGAATATG	TATTGCTGGC	ATTTTCGGTG	GCGGTGTTGT
	851	CCGTGTGCTG	CCTGTTTCCT	TTGTCGGCAA	TTGTTGTGAA	AGCGTGGTCG
40	901	CGCGCGGAAT	CGCGGCGGTG	GTTAATGGAA	AGTGAACGT	GGCAGGCAGT
	951	GTGGAATACT	ttGCGCTTTT	CGGCGGCGGC	GGTGTTCGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGGCGC	GGCTGGTGTG	GATGCGCGGA
	1051	CTGGTGTTTT	TACCGTTTAT	GGTGTCGCGG	GTTTGTGTTT	CGGCGGGCGT
	1101	GCTGCTGCTT	TATCCGGGGT	GGACGGCTTC	GTTACCGCTG	CTGCTGGCGA
45	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCGGCC
	1201	TGGGATGCAC	TGCCGCCGGA	TTACGGCAGG	GCGGCGGCAG	GTTTGGGCGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGTATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	CGGTCTGACT	TTGGCGGCGG	CGACGTGTGT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTCTGTGC	GCGTCCGGAA	TGGCAGACGT	TGACGACTTT
50	1401	GATTATATGCC	TATTTGGGGC	GTGCGGGTGA	GGACAATTAT	GCGCGGGCAA
	1451	TGGTGTGAC	ATTGCTGTTG	TCGGCATTTC	CGGTGTGCAT	TTTCTGCTG
	1501	TTGACAACG	GCGAAGGCGg	aaaACGGACG	GAAACGTTAT	AA

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

55	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLPPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPPT
	101	LVAGVGV <del>LAL</del>	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPLV	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSR <del>YAT</del>	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPPPQSV	GEYVLLAFSV	AVLSVCCLFP	LSAIVVKAWS
60	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVL <del>SA</del>
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
	501	LDN <del>EGG</del> GKRT	ETL*			

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

	orf139ng	MDGRCWAVRGAFSLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
	orf139-1	MDGRRWVVGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
5	orf139ng	ATCVLVLPPLGVFPAWVLARLAFPGRALVLRLLMLPFVMP TLVAGVGV LALFGADG LLWRG
	orf139-1	ATCVLVLPPLGVFPAWVLARLAFPGRALVLRLLMLPFVMP TLVAGVGV LALFGADG LLWRG
10	orf139ng	RQDTPYLLLYGNVFFNLPVLVRAAYQGFAQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
	orf139-1	RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
15	orf139ng	WLAGGVCLVFLYCFSGFGLALLGGSR YATVEVEIYQLVMFELDMAGASALVWLVLGVTA
	orf139-1	WLAGGVCLVFLYCFSGFGLALLGGSR YATVEVEIYQLVMFELDMAVASVLVWLVLGVTA
20	orf139ng	AAGLLYAWFGRRAVSDKAVSPVMPSP PQSVGEYVLLAFSVAVLSVCCLFPLSAIVVKAWS
	orf139-1	AAGLLYAWFGRRAVSDKAVSPVMPSP PQSVGEYVLLAFSAVLSVCCLFPLLAIVVKAWS
25	orf139ng	AGESRRVLMES ETWQAVWNTLRFSA AAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
	orf139	AGESRRVLMES ETWQAVWNTLRFSA AAVYAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
	orf139ng	VCVSAGVLLLYPGWTASLPLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
	orf139-1	VCVSAGVLLLYPGWTASLPLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
30	orf139ng	QTACRITFPLLPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
	orf139-1	QTACRITFPLLPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
35	orf139ng	ARAMVLTLLLSAFVAVCIFLLLDNGEGGKRTETL
	orf139-1	ARAMVLTLLLSAFALGIFLLLDGGEGGKQTETL

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCCGCG	CAAACCCTGT	TGGGCATTTT
45	51	GGCGGCGGCA	ATCATCCTCA	TTCTGATTTT	AATCGTCAGA	TTCCGCATCC
	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTGACGGC	TTTGGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	TGTCAAAGAC	ATACTGGTCA	AAAACCTTCGG
	201	CGGCACGCTC	GGCGGCGTGG	CGCTTCTGGT	CGGCCTGGGC	GCGATGCTCG
	251	AACGTTTGGT	C...			

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

50	1	MDGWTQTL	SAQTLLGISAAA	IILILILIVR	FRIHALLTLV	IVSLLTALAT
	51	GLPTGSIVKD	ILVNFGGTL	GGVALLVGLG	AMLERLV..	

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCCGCG	CAAACCCTGT	TGGGCATTTT
55	51	GGCGGCGGCA	ATCATCCTCA	TTCTGATTTT	AATCGTCAAA	TTCCGCATCC
	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTGACGGC	TTTGGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	TGTCAACGAC	ATACTGGTCA	AAAACCTTCGG

-334-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135

201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CCGCCTGGGC GCGATGCTCG  
251 GACGTTTGGT CGAAACATCC GCGGCGGCAC AGTCGCTGGC GGACGCGCTG  
301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC  
351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAACAGGA CGTACTGCC  
451 TTCGCGCTTG CCTCCATCGG CGCATTTCCT GTCATGCACG TCTTCCTGCC  
501 GCGCCATCCG GCGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCTCT  
651 TCCCGAATG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
751 ATTTTCCTGA ATACCGCGCT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
801 TGCGGACGAA ACCTGGGTTC AGACGGCAA AATAATCGGT TCGACACCGA  
851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA  
901 CCGCGCGAAA GCGGCGGCG GTTGGAAAA ACCGTGGACG GCGCACTCGC  
951 CCCCGTCTGT TCCGTGATTC TGATTACCGG CCGGGGCGGT ATGTTGCGCG  
1001 GCGTTTTGCG CGCTTCCGCG ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
1051 GATTTGGGCA TTCCCGTCCT TTTGGGCTGT TTCCTTGTCT CCTTGGCACT  
1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCTGACCACC GCGCGCGCGC  
1151 TGATGGCTCC TGCGGTTGCG GCGCGCGGT TTACCGACTG GCAGCTCGCC  
1201 TGTATCGTAT TGCAACGGC GGCAGGTTG GTCGGTTGCA GCCACTTCAA  
1251 CGACTCCGCG TTCTGGCTGG TCGGCCGTCT CTTGGACATG GACGTACCGA  
1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATCGCG  
1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT  
51 GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL  
101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP  
151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
201 SGYMLGKVLG RTIHVPPEL LSGCTQDNDL PKEPAKAGTV VAIMLIPMLL  
251 IFLNTGVSAL ISEKLVSADL TWVQTAIIIG STPIALLISV LVLEFVLGRK  
301 RGESESALEK TVDGALAPVC SVILITGAGG MFGGVLRASG IGKALADSMA  
351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
401 CIVLATAAGS VGCSEFNDSE FWLVGRLLDM DVPTTLKWT VNOTLIALIG  
451 FALSALLFAI V\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N.*

*meningitidis*:

40  
45  
50

orf140.pep	MDGWTQTLAQTLLGISAAAIIILILILIVFR	10	20	30	40	50	60
orf140a	MDGWTQTLAQTLLGISAAAIIILILILIVK	10	20	30	40	50	60
orf140.pep	ILVKNFGGTLGGVALLVGLGAMLERLV	70	80				
orf140a	VLVKNFGGTLGGVALLVGLGAMLRVETSGGAQSLADALIRMFGEKRAPFALGVASLIF	70	80	90	100	110	120

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

55  
60

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT  
51 GCGGCGGCGA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC  
101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC  
151 GGTTTGCCCA CAGGCAGCAT TGTCAACGAC GTACTGGTCA AAAACTTCGG  
201 CCGCACGCTC GCGGCGGTGG CGCTTCTGGT CCGCCTGGGC GCGATGCTCG  
251 GACGTTTGGT CGAAACATCC GCGGCGGCAC AGTCGCTGGC GGACGCGCTG  
301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC  
351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAACAGGA CGTACTGCC  
451 TTCGCGCTTG CCTCCATCGG CGCATTTCCT GTCATGCACG TCTTCCTGCC

-335-

5 501 GCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT  
 651 TCCCGAACTG CTCAGCGCGC GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTTCTCTGA ATACCGCGGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TCGGGACGAA ACCTGGGTTC AGACGGCAAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAA  
 901 CGCGGCGAAA GCGGCAGCGC GTTGGAAAAA ACCGTGGACG GCGCACTCGC  
 10 951 CCCCCTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGGC  
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCCT TTGGGCTGT TTCCTGTCTG CCTTGGCACT  
 1101 GCGTATCGCG CAAGGTCGG CAACCGTCGC CCTGACCACC GCCGCCGCGC  
 1151 TGATGGCTCC TGCCGTTGCC GCCCGCGGCT TTACCGACTG GCAGCTCGCC  
 15 1201 TGTATCGTAT TGGCAACGGC GGCAGGTCG GTCGGTTGCA GCCACTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCTT CTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAA CCCTCATCGC ACTCATCGGC  
 1351 TTTGCCTGT CCGCACTGCT GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT  
 51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRVLVETS GGAQSLADAL  
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP  
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL  
 25 251 IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK  
 301 RGESGSALEK TVDGALAPVC SVILITGAGG MEGGVLRASG IGKALADSMA  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSEHFNDG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG  
 451 FALSALLFAI V\*

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

20	orf140-1.pep	MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT GLPTGSIVND	60
	orf140a	MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT GLPTGSIVND	60
35	orf140-1.pep	ILVKNFGGTL GGVALLVGLG AMLGRVLVETS GGAQSLADAL IRMFGEKRAP FALGVASLIF	120
	orf140a	ILVKNFGGTL GGVALLVGLG AMLGRVLVETS GGAQSLADAL IRMFGEKRAP FALGVASLIF	120
40	orf140-1.pep	GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP FALASIGAFS VMHVFLPPHP GPIAASEFYG	180
	orf140a	GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP FALASIGAFS VMHVFLPPHP GPIAASEFYG	180
45	orf140-1.pep	ANIGQVLILG LPTAFITWYF SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV	240
	orf140a	ANIGQVLILG LPTAFITWYF SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV	240
50	orf140-1.pep	VAIMLIPMLL IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK	300
	orf140a	VAIMLIPMLL IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK	300
55	orf140-1.pep	RGESGSALEK TVDGALAPVC SVILITGAGG MEGGVLRASG IGKALADSMA DLGIPVLLGC	360
	orf140a	RGESGSALEK TVDGALAPVC SVILITGAGG MEGGVLRASG IGKALADSMA DLGIPVLLGC	360
60	orf140-1.pep	FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA CIVLATAAGS VGCSEHFNDG	420
	orf140a	FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA CIVLATAAGS VGCSEHFNDG	420
	orf140-1.pep	FWLVGRLLDM DVPTTLKTWT VNQTLIALIG FALSALLFAI	461
	orf140a	FWLVGRLLDM DVPTTLKTWT VNQTLIALIG FALSALLFAI	461

Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

```

orfl40.pep  MDGWTQTLSAQTLGISAIAIILILILIVFRIRALLTLVIVSLLTALATGLPTGSIVKD  60
            ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orfl40ng    MDGRTQTLSAQTLGISAIAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND  60

5  orfl40.pep  ILVKNFGGTLGGVALLVGLGAMLERLV  87
            :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orfl40ng    VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF  120

```

The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEQ ID 590>:

```

10      1  MDGRTQTLSA  QTLGISAIAA  IILILILIVK  FRIRALLTLV  IASLLTALAT
      51  GLPTGSIVND  VLVKNFGGTL  GGVALLVGLG  AMLGRLVETS  GGAQSLADAL
     101  IRMFGEKRAP  FAPGVASLIF  GFPIFFDAGL  IVMLPIVFAT  ARRMKQDVLP
     151  FALASVGAFS  VMHVFLPPHP  GPIAASEFYG  ANIGQVLLIG  LPTAFITWYF
     201  SGYMLGKVLG  RAIHVPVPEL  LSGGTQSDP  PKEPAKAGTV  VAVMLIPMLL
15      251  IFLNTGVSAI  ISEKLVSAD  TWVQTAKMIG  STPVALLISV  LAALLVLGRK
     301  RGESGSTLEK  TVDGALAPAC  SVILITGAGG  MFGGVLRASG  IGKALADSMA
     351  DLGIPVLLGC  FLVALALRIA  QGSATVALTT  AAALMAPAVA  AAGFTDWQLA
     401  CIVLATAAGS  VGCSHFND  FWLVGRLSDM  DVPTTLKTWT  VNQTIAFIG
     451  FALSALLFAI  V*

```

20 Further work revealed a variant gonococcal DNA sequence <SEQ ID 591>:

```

      1  ATGGACGGCC  GGACACAGAC  GCTGTCCGCG  CAAACCTTGT  TGGGCATTTG
      51  GCGCGCGGCA  ATCATCCTCA  TTCTGATTTT  AATCGTCAAA  TTCCGCATCC
     101  GCGCGCTGCT  GACACTGGTC  ATCGCCAGCC  TGCTGACGGC  TTTGGCAACC
25      151  GGTTCGCCCA  CAGGCAGCAT  CGTCAACGAC  GTACTGGTCA  AAAACTTCGG
     201  CGGCACGCTC  GCGCGCGTGG  CGCTTCTGGT  CGTCTGGGC  GCAATGCTCG
     251  GACGTTTGGT  AGAAACATCC  GCGCGCGCAC  AGTCGCTGGC  GGACGCGCTG
     301  ATCCGGATGT  TCGGCGAAAA  ACGCGCACCG  TTCGCTCCGG  GCGTTGCCTC
     351  GCTGATTTTC  GGCTTCCCGA  TTTTCTTCGA  TGCCGGACTA  ATCGTCATGC
     401  TGCCCATCGT  ATTCGCCACC  GCACGGCGCA  TGAACAGGA  CGTACTGCCC
30      451  TTCGCGCTTG  CCTCCGTCGG  CGCATTTTCC  GTCATGCACG  TCTTCCTGCC
     501  GCCCCATCGG  GGCCCGATTG  CCGCTTCCGA  ATTTTACGGC  GCGAACATCG
     551  GCCAGGTTTT  GATTTTGGGT  CTGCCGACCG  CCTTCATCAC  ATGGTATTTT
     601  AGCGGCTATA  TGCTCGGCAA  AGTGTGGGG  CGCGCCATCC  ATGTTCCCGT
35      651  TCCCGAAGTG  CTCAGCGGCG  GCACGCAAGA  CAGCGACCCG  CCGAAAGAAC
     701  CTGCCAAGC  AGGAACGGTC  GTCGCCGTC  TGCTGATTCC  CATGCTGCTG
     751  ATTTTCTGA  ATACCGCGGT  ATCAGCCCTC  ATCAGCGAAA  AACTCGTAAG
     801  TGCGGACGAA  ACTTGGGTTC  AGACGGCAA  AATGATCGGT  TCGACACTG
     851  TCGCCCTTCT  GATTTCCGTA  TTGGCCGCAC  TGTGGTCTT  GGGACGCAA
40      901  CCGGCGGAAA  GCGGCAGCAC  GTTGGAAAA  ACCGTGGACG  GCGCACTCGC
     951  CCCCGCCTGT  TCCGTGATTC  TGATTACCGG  CGCGGCGGT  ATGTTGCGG
    1001  GCGTTTTGCG  CGCTTCCGCG  ATCGGCAAGG  CACTCGCGCA  CAGCATGGCG
    1051  GATTTGGGCA  TTCCCGTCCT  TTTGGGCTGC  TTCCTTGTCG  CCTGGCACT
    1101  GCGTATCGCG  CAAGGTTCCG  CAACCGTCGC  CTGACCACA  GCCGCCGCGC
    1151  TGATGGCTCC  TGCCGTTGCG  GCCGCCGGCT  TTACCGACTG  GCAGCTCGCC
45      1201  TGTATCGTAT  TGCAACGGC  GGCAGGTTGC  GTCGGTTGCA  GCCACTTCAA
     1251  CGACTCCGGC  TTCTGGCTGG  TCGGCCGCCT  CTTGGATATG  GACGTACCGA
     1301  CCACGCTGAA  AACTGGACG  GTCAACCAA  CCCTCATCGC  ATTCATCGGC
     1351  TTTGCCTTGT  CCGCACTGCT  GTTTGCCATC  GTCTGA

```

This corresponds to the amino acid sequence <SEQ ID 592; ORF140ng-1>:

```

50      1  MDGRTQTLSA  QTLGISAIAA  IILILILIVK  FRIRALLTLV  IASLLTALAT
      51  GLPTGSIVND  VLVKNFGGTL  GGVALLVGLG  AMLGRLVETS  GGAQSLADAL
     101  IRMFGEKRAP  FAPGVASLIF  GFPIFFDAGL  IVMLPIVFAT  ARRMKQDVLP
     151  FALASVGAFS  VMHVFLPPHP  GPIAASEFYG  ANIGQVLLIG  LPTAFITWYF
55      201  SGYMLGKVLG  RAIHVPVPEL  LSGGTQSDP  PKEPAKAGTV  VAVMLIPMLL
     251  IFLNTGVSAI  ISEKLVSAD  TWVQTAKMIG  STPVALLISV  LAALLVLGRK
     301  RGESGSTLEK  TVDGALAPAC  SVILITGAGG  MFGGVLRASG  IGKALADSMA
     351  DLGIPVLLGC  FLVALALRIA  QGSATVALTT  AAALMAPAVA  AAGFTDWQLA
     401  CIVLATAAGS  VGCSHFND  FWLVGRLSDM  DVPTTLKTWT  VNQTIAFIG
     451  FALSALLFAI  V*

```

60 ORF140ng-1 and ORF140-1 show 96.3% identity over 461aa overlap:

```

orfl40ng-1.pep MDGRTQTLSAQTLGISAIAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND
                ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

-337-

```

5 orf140-1      MDGWTQTLAQTLGISAIAIILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
  orf140ng-1.pep VLKKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
  orf140-1      ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
10 orf140ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASVGAFSVMHVFLPPHPGPIAASEFYG
  orf140-1      GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASIGAFSVMHVFLPPHPGPIAASEFYG
  orf140ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQSDPDPKEPAKAGTV
  orf140-1      ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPVPELLSGGTQDNDLPKEPAKAGTV
15 orf140ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKMIGSTPVALLISVLAALLVLGRK
  orf140-1      VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK
20 orf140ng-1.pep RGESGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
  orf140-1      RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
  orf140ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
25 orf140-1      FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
  orf140ng-1.pep FWLVGRLLDMDVPTTLKTTWTVNQTLLIAFIGFALSALLFAIV
  orf140-1      FWLVGRLLDMDVPTTLKTTWTVNQTLLIALIGFALSALLFAIV

```

30 Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

```

35 gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
   This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
   protein GNTP BACLI SW: P46832 [Escherichia coli] Length = 454
   Score = 210 bits (529), Expect = 1e-53
   Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

40 Query: 88  ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
   E SGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
   Sbjet: 80  EHSAGAESLANYFSRKLGDKRTIAALTIAAFFLGIPVFFDVGFILAPIIYGFYAKVAKIS 139

   Query: 148 VLPPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
   L F L G +HV +PPHPG+AA+ A+IG + I+G+ + I GY K
   Sbjet: 140 PLKFGPLVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVGVVGYFAAK 198

45 Query: 208 VLGRAIHVPVPELL-----SGGTQSDPDPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
   ++ + + E+L G T+ SD P A V +++IP+ +I T
   Sbjet: 199 IINKRQYAMSVEVLEQMLAPASEEGATKLSDKINPPGVA-LVTSILIVIPAIIMAGT-- 255

50 Query: 258 SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXGRKRGESGSTLEKTVDGALA 317
   +S L+ + T ++IGS +RG S + AL
   Sbjet: 256 ---VSATLMPPSHPLLGLTQLIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312

   Query: 318 PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXX 377
   A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS
55 Sbjet: 313 TAAVVILVTGAGGVFGKVLVESGVGKALANMLQMLDPLLPAAFIISLALRASQGS--AT 370

   Query: 378 XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFNDSGFWLVGRLLDMDVPTTLK 437
   G Q + LA G +G SH NDSGFV+V + L + V LK
60 Sbjet: 371 VAILTTGGLLSEAVMGLNFIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGLSVADGLK 430

   Query: 438 TWTVNQTLLIAFIGFALSALLFAIV 461
   TWTV T++ F GF ++ ++A++
   Sbjet: 431 TWTVLTTILGFTGFLITWCVWAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence  
 65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
      51  TTTGCTGTCT CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
     101  GCGTATTTTT TGCCGTTATC GGACTGACTT CCTGCGGCTT TGCCGGTTTC
     151  AACTTTTTTG GCAGACACCA CGGGCGCAC. GTCGTCTCTGA TTCTCATCGG
     201  CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCCG
    10  251  CCTTGCCGCG CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCCG
     301  CGCGTGATTG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
     351  GTTGGCAGCA GCTTATCCGG CAGCATTGTC CCTGATGCTG CCCTTGCCCC
     401  TACTGATGTT TTTCCGTCCG ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1  ..DFGISPVYLW VAAAFKHLIS PWAADSYDVA RFAGVFFFAVI GLTSCGFAGF
     51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAFAAAGL VLHGYSLARR
     101  RVIAASFLLG TGWTLMSLAA AYPAAFALML PLPVLMFRRP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1  ATGCTGACCT ATACCCCGCC CGATGCCC GC CCGCCCGCCA AAACCCACGA
     51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
    101  TGTTTTCCCA CGATTGTGGG AATCCTGACG AACCTGCCGT CTATACCGCC
    151  GTCGAAGCAC TGGCAGGCAG CCCCACCCCG TTGGTTGCCG ATCTGTTCGG
    201  TCAAACCGAT TTCGGCATA CCGCCGTGTA TCTTTGGGTT GCCGCGCGCT
    251  TCAAACATTT GCTGTGCGCG TGGGCTGCCG ACTCATACGA TGCCGCACGC
    301  TTTGCGAGCG TATTTTGTGC CGTTATCGGA CTGACTTCCT GCGGCTTTGC
    351  CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
    401  TCATCGGCTG TATCGGGCTG ATTCCAGTTG CCCATTTCCT CAACCCCGCT
    451  GCGCGCGCCT TTGCCCGCGC CGGACTGGTG CTGCACGGTT ATTCTTTGGC
    501  TCGCCCGGCG GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
    551  TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTTGCCCT GATGCTGCCC
    601  TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
    651  GACGGCAGTC GCCTCACTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
    701  CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTTCGCGCA ATGGCTCGAC
    751  TATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACGTTT AGACGGCATT
    801  CAGTTTGTGT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC
    851  TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC
    901  TGCGGGGATT TGGGCGTCTG CTGGATGCTT GCCGTTTGGT TGCTGCTTGC
    951  CGTCAATCCG CAGCGTTTTC AGGATAACCT CGTCTGGCTG CTTCCGCGCG
   1001  TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGCGC CGGCGCGGCG
   1051  GCGTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGAAGT GTTCCGTGTT
   1101  CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
   1151  CCGAACGCGC CGCCTATTTT AGCCCGTATT ATGTTCTCTGA TATCGATCCC
   1201  ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTGGC TGTGGGCGAT
   1251  TACCCGAAAA AACATACCGC GCAGGCAGGC GGTACCAAC TGGGCGGCAG
   1301  CGGTTACCCCT GACCTGGGCT TTGCTGATGA CGCTGTTTCT GCCGTGGCTG
   1351  GACGCGGCGA AAAGCCACGC GCCGTCGTC CGGAGTATGG AGGCATCGCT
   1401  TTCCCCGGA TTTGAAACGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
   1451  TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
   1501  TTGCCGCACC GCGTCGGCGA TGTACAATGC CGTACCGCA TCGTCTCTCT
   1551  GCCCCAAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
   1601  CGCGTCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAATCGGG
   1651  GAAAATATAT AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1  MLTYTPPDAR PPAKTHEKFW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
     51  VEALAGSPTF LVAHLFGQTD FGIPPVYLWV AAFAKHLSP WAADSYDAAR
    101  FAGVFFFAVIG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL LIPVAHFLNPA
    151  AAAFAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
    201  LPVLMFRRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD

```



-339-

251 YHVFQTFGGV RHVQTAFLSF YYLKNLLWFA LPALPLAVWT VCRTRLFSTD  
 301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDLRRGAA  
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP  
 401 IPMAVAVLFT PLWLWAIKTRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL  
 451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIGIGGGDLH TRIVWTOYGT  
 501 LPHRVGDVQC RYRIVLLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG  
 551 ENI\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

```

                                10      20      30
orf141.pep                      DFGISPVYLWVAAAFKHLSPWAADSYDVA
15 orf141a      WNPDEPAVYTAVEALAGSPTPLVAHLFGQIDFGIPVYLWVAAAFKHLSPWAADPYDAA
                   40      50      60      70      80      90

                                40      50      60      70      80      90
orf141.pep      RFAGVFFFAVIGLTSFCGFAFENFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL
20 orf141a      RFAGVFFFAVIGLTSFCGFAFENFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL
                   100     110     120     130     140     150

                                100     110     120     130     140
orf141.pep      VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP
25 orf141a      VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRMLMTA
                   160     170     180     190     200     210

30 orf141a      VASLAFALPLMTVYPLLLAKTQPALFAQWLDHVFQTFGGVRHIQTAFSLFYLLKNLLWF
                   220     230     240     250     260     270

```

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

```

1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
35 51  AAAGCCGTGG CTGTTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
101 TGTTTTCCCA CGATTGTGTT AATCCTGACG AACCTGCCGT CTATACCGCC
151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCT TTGGTTGCCC ATCTGTCGG
201 TCAAATCGAT TTCGGCATA CCGCCGTGTA TCTTTGGGTT GCCGCGCGT
251 TCAAACATTT GCTGTGCGCG TGGGCTGCCG ACCCGTATGA TGCCGACGCG
301 TTTGCCGCGC GTTTTTTCGC CGTTGTCGGA CTGACTTCCT GCGGCTTTGC
40 351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
401 TCATCGGCTG TATCGGGCTG ATTCGACCG TACACTTTCT CAACCCGCT
451 GCCGCGCCTT TTGCCCGCG CCGACTGGTG CTGCACGGTT ATTCTTTGGC
501 TCGCCGGCGC GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGTGGACGCG
551 TGATGTCGTT GGCAGCAGCT TATCCGCGCG CATTTGCCCT GATGCTGCC
45 601 CTGCCCGTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
651 GACGCGAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTATG ACCGTTTACC
701 CGCTGCTCTT GGCAAAAACG CAGCCGCGCG TGTTCCGCGA ATGGCTCGAC
751 GATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACATTC AGACGCGATT
801 CAGTTTGTGT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCTGCGC
50 851 TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCTGTTT TTCGACCGAC
901 TGGGGGATTT TGGGCGTGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC
951 CGTCAATCCG CAGCGTTTTC AGGATAACCT CGTCTGGCTG CTTCCGCGCG
1001 TTGCCCTGTT CGGCGCGCGC CAACTGGACA GCCTGAGACG CGGCGCGGCG
1051 GCGTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGAAGTT TTGCCGTGTT
55 1101 CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
1151 CCGAACGCGC GCGCTATTTC AGCCGTATT ATGTTCTTGA TATCGATCCC
1201 ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTGGC TGTGGGCGAT
1251 TACCCGCAAA AACATACGCG GCAGGCAGGC GGTACCAAC TGGGCGGCG
1301 GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
60 1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGCT
1401 TTCCCGGAA TTAACCGGG AGCTTTCAGA CGGCATCGAG TGTATCGACA
1451 TAGGCGGCGG CGACCTACAC ACGCGGATT TTTGGACGCA GTACGCGACA
1501 TTGCCGACAC GCGTCGCGCA TGTACAATGC CGTACCGCA TCGTCCGCTT
1551 GCCCAAAAC GCGGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGCTG

```

-340-

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG  
 1651 GAAATATAT TAAAAACAAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

5 1 MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA  
 51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLLSP WAADPYDAAR  
 101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA  
 151 AAFAAGLV LHGYSLARRR VIAASELLGT GWTLMSLAAA YPAAFALMLP  
 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD  
 251 DHVFGTFGGV RHIQTAFSLF YYLKNLLWEA LPALPLAVWT VCRTRLFSTD  
 10 301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDSLRRGAA  
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYYVPDIDP  
 401 IPMAVAVLFT PLWLWAI TRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL  
 451 DAAKSHAPV RSMEASLSPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT  
 501 LPHRVGDVQC RYRIVRLPQN ADAPQGWQTV WQGAPRNKD SKFALIRKTG  
 15 551 ENILKTTD\*

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

orfl41a.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP  
 20 orfl41-1 MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP  
 orfl41a.pep LVAHLFGQIDFGIPPVYLWVAAFAKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN  
 orfl41-1 LVAHLFGQTDGIPPVYLWVAAFAKHLLSPWAADSYDAARFAGVFFAVIGLTSFGFAGFN  
 25 orfl41a.pep FLGRHHGRSVVLILIGCIGLIPTVHFLNPAFAAAGLVLHGYSLARRRVIAASELLGT  
 orfl41-1 FLGRHHGRSVVLILIGCIGLIPVAHFLNPAFAAAGLVLHGYSLARRRVIAASELLGT  
 orfl41a.pep GWTLMSLAAAYPAAFALMLPLPVLMEFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT  
 30 orfl41-1 GWTLMSLAAAYPAAFALMLPLPVLMEFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT  
 orfl41a.pep QPALFAQWLDHVFSTFGGVRHIQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD  
 35 orfl41-1 QPALFAQWLDYHVFSTFGGVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD  
 orfl41a.pep WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 40 orfl41-1 WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 orfl41a.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMMAVAVLFTPLWLWAI TRK  
 orfl41-1 FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMMAVAVLFTPLWLWAI TRK  
 45 orfl41a.pep NIRGRQAVTNWAAGVTLTALLMTLFLPWLDAKSHAPVRSMEASLSPELKRELSDGIE  
 orfl41-1 NIRGRQAVTNWAAGVTLTALLMTLFLPWLDAKSHAPVRSMEASLSPELKRELSDGIE  
 orfl41a.pep CIDIGGGDLHTRIVWTQYGTLPFRVGDVQCRYRIVRLPQNADAPQGWQTVWQGARPRNKD  
 50 orfl41-1 CIGIGGGDLHTRIVWTQYGTLPFRVGDVQCRYRIVRLPQNADAPQGWQTVWQGARPRNKD  
 orfl41a.pep SKFALIRKTGENI  
 55 orfl41-1 SKFALIRKIGENI

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

60 orfl41.pep DFGISPVYLWVAAFAKHLLSPWAADSYDVA 30  
 orfl41ng WNPAEPVYTAVEALAGSPTPLVAHLFGQTDGIPPVYLWVAAFAKHLLSPWAADHPYDAA 126

5	orf141.pep	RFAGVFFAVIGLTS	CGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL	90
	orf141.ng	RFAGVFFAVIGLTS	CGFAGFNFLGRHHGRSVVLIHIGCIGLIPVAHFFNPAAAAFAAAGL	186
5	orf141.pep	VLHGYSLARRRVIAAS	FLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP	140
	orf141.ng	VLHGYSLARRRVIAAS	FLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRMLMTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10	1	MPSEAVSARP	LCEYLLHLAI	RPFLTLMLT	YTPPDARPPA	KTHEKPWLLL
	51	LMAFAWLWPG	VFSDHLWNPA	EPAVYTAVEA	LAGSPTPLVA	HLFGQTDGFI
	101	PPVYLWVAAA	FKHLLSPWAA	HPYDAARFAG	VFFAVIGLTS	CGFAGFNFLG
	151	RHHGRSVVLI	HIGCIGLIPV	AHFFNPAAAA	FAAAGLVLHG	YSLARRRVIA
15	201	ASFLLGTGWT	LMSLAAAYPA	AFALMLPLPV	LMFFRPWQSR	RLMLTAVASL
	251	AFALPLMTVY	PLLLAKTQPA	LFAQWLNHYV	FGTFGGVRHI	QRAFSLFHYL
	301	KNLLWFAPPG	LPLAVWTVCR	TRLFSTDWGI	LGIVWMLAVL	VLLAFNPQRF
	351	QDNLVWLLPP	LALFGAAQLD	SLRRGAAAFV	NWFGIMAFGL	FAVFLWTGFF
20	401	AMNYGWPAKL	AERAAFYSPY	YVPDIDPIPM	AVAVLFTPLW	LWAITRKNIR
	451	GRQAVTNWAA	GVTLTWALLM	TLFLPWLDAA	KSHAPVVRSM	EASFSPLEKR
	501	ELSDGIECIG	IGGGDLHTRI	VWTQYGTLPH	RVGDVRCRYR	IVRLPQNADA
	551	PQGWQTVWQG	ARPRNKDSKF	ALIRKIGENI	LKTTD*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

25	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
	51	AAAACCGTGG	CTGCTGCTGT	TGATGGCGTT	TGCCTGGCTG	TGGCCCGGCG
	101	TGTTTTCCCA	CGATTTGTGG	AATCCTGCCG	AACCTGCCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCC	TTGGTTGCCC	ATCTGTTCCG
30	201	TCAAACCGAT	TTCGGCATA	CGCCCGTGTA	TCTTTGGGTT	GCCGCCGCAT
	251	TCAAACATTT	GCTGTCGCCG	TGGGCAGCCG	ACCCGTATGA	TGCCGCACGC
	301	TGTGCAGCG	TATTTTTTGC	CGTTATCGGA	CTGACTTCTT	GCGGCTTTGC
	351	CGGTTTCAAC	TTTTTGGGCA	GACACCACGG	GCGCAGCGTT	GTTTTAATCC
35	401	ATATCGGCTG	TATCGGGCTG	ATTCCGGTTG	CCCATTTCCT	CAATCCcgcc
	451	gcccgcgcct	tTGCCCGCCG	CGGACTGGTG	CTGCacggct	actcgcgtgGC
	501	AGCCCGGCGC	GTGATTgcgc	cctctTtccT	GCTCGGTACG	GGTTGGACGT
	551	TGATGTCGCT	GGCGGCAGCT	TATCCGCGCG	CGTTTGCCT	GATGCTGCCC
40	601	CTGCCCCGTG	TGATGTTTTT	CCGTCCGTGG	CAAAGCAGGC	GTTTGATGTT
	651	GACGGCAGTC	GCCTCGCTTG	CCTTTGCCCT	GCCGCTTATG	ACCGTTTACC
	701	CGCTGCTCtt	gGCAAAAACG	CAGCCCGCGC	TGTTTGCGCA	ATGGCTCAAC
	751	TATCACGTTT	TCGGTACGTt	cggcgGCGTG	CGGCACaTTC	AGAggGCatT
45	801	Cagttttgttt	cactatctgA	AAaatctgct	ttggttcgca	ccgccccggcC
	851	TGCCGCTGGC	GGTTTGGACG	GTTTGCCGCA	CAGCCTGTT	TCGACCGAC
	901	TGGGGGATTT	TGGGCATTGT	CTGGATGCTT	GCCGTTTGG	TGCTGCTCGC
	951	CTTTAATCCG	CAGCGTTTTC	AAGACAACCT	CGTCTGGCTG	CTGCCCGCGC
50	1001	TGCCCCGTGT	CGGCGCGGCG	CAACTGGACA	GCCTGAGGCG	CGGCGCGGCG
	1051	GCTTTTGTCA	ACTGGTTTCG	CATTATGGCG	TTCGGGCTGT	TGCGCGTGT
	1101	CCTGTGGACG	GGCTTTTTCG	CCATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CGGAACGCGC	CGCTACTTTC	AGCCCGTATT	ACGTTCCCGA	CATCGATCCC
55	1201	ATTCCGATGG	CGGTTGCCGT	ACTGTTTACA	CCCTTGTTGC	TGTGGGCGAT
	1251	TACCCGGAAA	AACATACGCG	GCAGGCAGGC	GGTTACCAAC	TGGGCGGCAG
	1301	GCGTTACCCT	GACCTGGGCT	TTGCTGATGA	CGCTGTTTCT	GCCGTGGCTG
	1351	GACGCGGCGA	AAAGCCACGC	GCCCGTCGTC	CGGAGTATGG	AGGCATCGTT
60	1401	TTCCCCGGAA	TTAAACCGGG	AGCTTTCAGA	CGGCATCGAG	TGTATCGGCA
	1451	TAGGCGGCGG	CGACCTGCAC	ACGCGGATTG	TTTGACGCA	GTACGGCACA
	1501	TTGCCGCACC	GCGTCGGCGA	TGTCCGTTGC	CGCTACCGTA	TCGTCCGCCCT
	1551	GCCCCAAAAC	GCGGATGCGC	CGCAAGGCTG	GCAGACGGTC	TGGCAGGGTG
55	1601	CGCGCCCGCG	CAACAAAGAC	AGTAAGTTTG	CACTGATACG	GAAAATCGGG
	1651	GAAAATATAT	TAAAAACAAC	AGATTGA		

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

60	1	MLTYTPPDAR	FPAKTHEKPW	LLLLMAFAWL	WPGVFSDHLW	NPAEPAVYTA
	51	VEALAGSPTP	LVAHLFGQTD	FGIPPVYLWV	AAAFKHLLSP	WAADPYDAAR
	101	FAGVFFAVIG	LTSCGFAGFN	FLGRHHGRSV	VLIHIGCIGL	IPVAHFLNPA
	151	AAAFAAAGLV	LHGYSLARRR	VIAASFLG	WTLMSLAAA	YPAAAFALMLP
60	201	LPVLMFFREW	QSRRLMLTAV	ASLAFALPLM	TVYPLLLAKT	QPALFAQWLN
	251	YHVFGTFGGV	RHIQRAFSLF	HYLKNLLWFA	PPGLPLAVWT	VCRTRLFSTD
	301	WGILGIVWML	AVLVLLAFNP	QRFQDNLVWL	LPPLALFGAA	QLDSLRRGAA

-342-

351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP  
 401 IPMAVAVLFT PLWLWATR K NIRGRQAVTN WAAGVTLTWA LLMTLFLPLWL  
 451 DAAKSHAPVV RSMEASF SPE LKRELS DGIE CIGIGGGDLH TRIVWTQYGT  
 501 LPHRVGDVRC RYRIVRLPQN ADAPQGWQTV WQGAPRPNKD SKFALIRKIG  
 551 ENILKTTD\*

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

orf141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP  
 orf141-1 MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP  
 orf141ng-1.pep LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVIGLTSCGFAGFN  
 orf141-1 LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTSCGFAGFN  
 orf141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAGLVHGYSLARRRVIAASFLLG  
 orf141-1 FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAGLVHGYSLARRRVIAASFLLG  
 orf141ng-1.pep GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT  
 orf141-1 GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT  
 orf141ng-1.pep QPALFAQWLNHYHVFETFGGVRHIQRAFSLFHYLKNLLWFAPPGLPLAVWTVCRTRLFSTD  
 orf141-1 QPALFAQWLDYHVFETFGGVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD  
 orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 orf141-1 WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMVAVLFTPLWLWATR K  
 orf141-1 FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMVAVLFTPLWLWATR K  
 orf141ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPLWLDAAKSHAPVRSMEASFPELKRELS DGIE  
 orf141-1 NIRGRQAVTNWAAGVTLTWALLMTLFLPLWLDAAKSHAPVRSMEASLSPELKRELS DGIE  
 orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGWQTVWQGAPRPNKD  
 orf141-1 CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVLLPQNADAPQGWQTVWQGAPRPNKD  
 orf141ng-1.pep SKFALIRKIGENILKTTDX  
 orf141-1 SKFALIRKIGENIX

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 72

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

1 ..CAATCCGCCA AATGGTTATC GGGCCAACT CTAGTCGGCA CAGCAATTGG  
 51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA  
 101 CCGGCCGCGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA  
 151 AGCGGTTTTC AGGTAGGCTA TACGTTTAA

This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

1 ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA  
 51 SGFQVGYTF\*

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

-343-

1 ATGGATAATT CGGGTAGTGA GGCGACAGGA AAATACCAAG GAAATATCAC  
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTC TATGTAAATT  
 101 ATGGACGTTT GATTGGCGGT ACGCCCGATG AGGAAAGTTT TGACGGCCAT  
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT  
 201 CGGTAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG  
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT  
 301 ACTGATTTTC GCTTCAACCG CCTGTGTAT CGTGATGCCA AACGCAAAAC  
 351 CTATCTCGGT GTAAAACGTG GGATGAGGGA AACAAAAACT TACATTGATG  
 401 ATGCCGAAC GACTGTACAA CGGCGTAAAA CTGCGGGTTG GTTGGCAGAA  
 10 451 CTTTCCACAA AAGAAATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA  
 501 ATATAAACGC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG  
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT  
 601 GTAAATACTC CTTTTCAAAAT CGGTAAACAG CTATTTGCCT ATGACACATC  
 651 CGTTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG  
 15 701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG  
 751 TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA  
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC  
 851 AATCCGCCAA ATGGTTATCG GGCCAACTC TAGTCGGCAC AGCAATTGGG  
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC  
 20 951 CGGCCGCGCA TTGAAAAAGC CCGAATTTT CCAATCAAGG AAATGGGCAA  
 1001 GCGGTTTTCA GGTAGCTAT ACGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH  
 51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAUSGLSE VYDYNKSYN  
 101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE  
 151 LSHKEYIGRS TADFKLKYKR GTGMKDALRA PEEAFGEGTS RMKIWTASAD  
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL  
 251 SAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAIG  
 301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF\*

30 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

35 orf142.pep QSAKWLSGQTLVGTAIGIRGQIKLGGNLHY 30  
 orf142ng RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY 313  
 orf142.pep DIFTGRALKKPEFFQSRKWASGFQVGYTF 59  
 40 orf142ng DIFTGRALKKPEYFQTKKWVTGFQVGYSF 342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTGA GGCGACAGGA AAATACCAAG GAAATATCAC  
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTC TATGTAAATT  
 45 101 ATGGACGTTT AATTGGCGGT ACGCCCGATG AGGAAATTTT TGACGGCCAT  
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT  
 201 CGGTAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG  
 251 CGGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAC  
 301 ACTGATTTTC GCTTCAACCG CCTGTGTAT CGTGATGCCA AACGCAAAAC  
 351 CTATCTCAGT GTAAAACGTG GGACGAGGGA AACAAAAAGT TACATTGATG  
 50 401 ATGCCGAAC GACTGTACAA CGGCGTAAAA CCACAGGTTG GTTGGCAGAA  
 451 CTTTCCACAA AAGGATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA  
 501 ATATAAACAC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG  
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT  
 601 GTAAATACTC CTTTTCAAAAT CGGTAAACAG CTATTTGCCT ATGACACATC  
 55 651 CGTTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG  
 701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG  
 751 CCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA  
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC  
 851 AATCCGCCAA ATGGTTATCG GGCCAACTC TAGCCGGCAC AGCAATTGGG  
 60 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC  
 951 CGGCCGTGCA TTGAAAAAGC CCGAATATTT TCAGACGAAG AAATGGGTAA

5

1	MDNSGSEATG	KYQGNITFSA	DNPFLSDMF	YVNYGRSIGG	TPDEENFDGH
51	RKEGSGNNYA	VHYSAFPFGKW	TWAFNHNNGYR	YHQAVSGLSE	VYDYNGKSKYN
101	TDFEGNRLLY	RDARKRTYLS	VKLWTRNETKS	YIDDAELTVQ	RKRTTGWLAS
151	LSHKGYIGRS	TADFLLKYKH	GTMKMDALRA	PEEAFGEFST	RMKIWTASAD
201	VPFPFQIGIK	LFAYDTSVHA	QWNKTPLTST	DKLAIGGHTH	VGTDFGEMSL
251	PAERGWYWRN	DLSQWFKPGH	QLYLQADVGH	VSGQSAKWL	QGTLAGTAIG
301	IRGQIKLGGN	LHYDIFTGRA	LKKPEYFQTK	KWVTGFQVGY	SF*

10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

15

```

orf142-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
|||||:|||||:|||||
orf142ng-1 MDNSGSEATGKYQGNITFSADNPFLGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA
|||||:|||||:|||||
orf142-1.pep VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYGKSYNTDFGFNRLLYRDAKRKYLG
|||||:|||||:|||||
orf142ng-1 VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYGKSYNTDFGFNRLLYRDAKRKYLS
|||||:|||||:|||||
orf142-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLKYKRGTMKDALRA
|||||:|||||:|||||
orf142ng-1 VKLWTRKTSYIDDAELTVQRRKTTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRA
|||||:|||||:|||||
orf142-1.pep PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKPTLTSQDKLAIGGHT
|||||:|||||:|||||
orf142ng-1 PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKPTLTSQDKLAIGGHT
|||||:|||||:|||||
orf142-1.pep VRGFDGEMSLSAERGMYWRNDSLWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLVGTAIG
|||||:|||||:|||||
orf142ng-1 VRGFDGEMSLPAERGMYWRNDSLWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIG
|||||:|||||:|||||
orf142-1.pep IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFQVGYTF
|||||:|||||:|||||
orf142ng-1 IRGQIKLGGNLHYDIFTGRALKKPEYFOTKKWVTGFOVGYSE
|||||:|||||:|||||

```

In addition, ORF142ng is homologous to the HecB protein of *E.chrysanthemi*:

40

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558  
Score = 119 bits (295), Expect = 3e-26  
Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)

45

Query: 2 DNSGSEATGKYQGNIIFSADNPFLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61  
 DNSG ++TG+ Q N + + DN FGL+D ++++ G S + + D + G  
 Subject: 230 DNSGOKSTGEEQLNGSLALDNVFLADOWFISAGHS---SRFATSHDAESLOAG----- 280

50

Query: 122 KLWTRKTSYIDDAELTVQRKTTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRAP 181  
R ++ + L RK + ++H + A F Y G +  
Sbjct: 340 TFSQRTGNNYLNGLSPSSSRKLSSVSLGVNHSOKLWGGGLATFNPTYNRGVRWLGSSETDT 399

55

Query: 182 EEAFGGETSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKPTLTSQDKLAIGGHTV 241  
 +++ E + WT SA P Y S++ Q++ L ++L +GG ++  
 Sbjct: 400 DKSADEPRAEFNKWTLASYYHPV---TDSITYLGSLYGOYSARALYGSEQLTLGGESSI 456

60

Query: 242 RGF<sup>+</sup>DGEMSLPAERGWYWRNDLSWQFKP----GHQLYLGA-DVGHSVGSQSAKWL<sup>+</sup>SGQTLAG 296  
 RGF E RG YWRN+L+WQ G+ ++ A D GH+ + +L G  
 Sbjct: 457 RGF-REQYTSNGRGAYWRNELNWOAWOLPVLGNVTFMAAVDGGHLYNHKODNSTAASLWG 515

Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGFGVGYSF 342  
 A+G+ + L + G + P + O V G++VG SF

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 609>:

This corresponds to the amino acid sequence <SEQ ID 610; ORF143>:

Further work revealed the complete nucleotide sequence <SEQ ID 611>:

This corresponds to the amino acid sequence <SEQ ID 612; ORF143-1>:

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF143 shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) from strain A of *N. meningitidis*:

[illegible]

-346-

```

orf143a      YGDENGHS DGINLSDEQLPLLM EQLSGSGKALLVDRNGLY LANANFHH EAAEELGLLAAE
              80          90          100          110          120          130

              100          110
5  orf143.pep  VAQMEKKYRLLIKNN
              |||||
orf143a      VAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFFLYIGSTK FILVIGGIPDLGKEA
              140          150          160          170          180          190

```

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

```

10      1  ATGGAATCAA CANTTTCACT ACAAGCAAAT TTATATCNCC GCCTGACTCC
        51  TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGNCCCCAGT GCCCGTAAAA
       101  CTTTGTGCA CAGCCTGTTG AAAGCGGATG CGGACGAAAT GGTNAGCAGT
       151  GAGAAGCTGC TTACCTGGGC GGANACCGCC GACATCGATA CCGCTTTGAA
       201  CCTGTTGTAC CGTTTGCAAA AACTCGAATT CCTCTATGGC GATGAAAACG
       251  GTCATTGAGA CGGCATCAAT TTGTCGGACG AGCAATGACC GTTGCTGATG
       301  GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
       351  GTATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT
       401  TGTGCGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCNNATT
       451  AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC
       501  CGGTCAGAGC GAATTGACAT TTTTCCCAT TATATCGGT TCAACCAAAT
       551  TTATTTGGT TATCGCGCGC ATTCCCGATT TGGGCAAAGA GGCATTGTGT
       601  ACTTTGGTAA GGATNTTATA CCNCCNGTTA CAGCAACCGC GTGTAAAACT
       651  TGGGAGAGAG GANGGTTAT GCAGCAATTA TTGA

```

This encodes a protein having amino acid sequence <SEQ ID 614>:

```

25      1  MESTXSLQAN LYXRLTPAGA FYAVSSDXPS AGKTLHSL LKADADEMVS S
        51  EKLLTWAXTA DIDTALNLLY RLQKLEFLYG DENGHS DGIN LSDEQLPLLM
       101  EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLXI
       151  KNNLYINNNA WGVCDPSGQS ELTFFFLYIG STKFILVIGG IPDLGKEAFV
       201  TLVRXLYXXL QQPRVKLGRE XGLCSNY*

```

30 ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

```

orf143a.pep  MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHSL LKADADEMVSSEKLLTWAXTA
orf143-1      MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLHSL LKADADEMVSSEKLLTWADTA

35  orf143a.pep  DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLM EQLSGSGKALLVDRNGLYLA
orf143-1      DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLM EQLSGSGKALLVDRNGLYLA

orf143a.pep  NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFFLYIG
40  orf143-1      NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFFLYIG

orf143a.pep  STKFILVIGGIPDLGKEAFVTLVRXLY
45  orf143-1      STKFILVIGGIPDLGKEAFVTLVRILY

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

```

50  orf143.pep  MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLXDEQLPLLM EQL 60
orf143ng      MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLM EQL 60

orf143.pep  SGSGKALLVDRNGLY LANANFHH EAAEELGLLAAEVAQMEKKYRLLIKNN 110
55  orf143ng      SGSGKALLVDRNGLY LANANFHH ESAEELGLLAAEVAQMEKKYRLIRNNLYINNNAWGV 120

```

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:



Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

15

20

25

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

30

35

40

45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

50

55

1	ATGACCTTTT	TACAACGTTT	GCAAGGTTTG	GCAGACAATA	AAATCTGTGC
51	GTTTGCATGG	TTCGTCGTCC	GCCGCTTTGA	TGAAGAACGC	GTACCCGCAGr
101	CGGCGGCAAG	CATGACGTTT	ACGACGCTTG	TGGCATTCTGT	CCCCGTGCTG
151	ACCGTGCATG	TGGCGCTCGC	TTCGATTTTC	CCCGTGTTTC	ACCGCTGGTC
201	GGATTGCTTC	GTCPTCTTCG	TCAACCAAAAC	CATTGTGCGC	CA. GGC CGGG
251	ACATGGTGTT	CGACTATATC	AATGCGTTCC	GCGAGCAGGC	GAACCGGCTG
301	ACGGCAATCG	GACGCGTGAT	GCTGGTCGTT	ACCTCGTGTA	TGCTGATTTCG
351	CACGATAGAC	AATACGTTCA	ACCGCATCTG	G <sub>u</sub> CGGGTCAA	wTyCCAGCGT
401	CCGTGGATG	..			

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLQRLQGL ADNKICAF AW FVVRRFDEER VPQXAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLLV TSLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTGCG TCGATTTTC CCCGTGTTCG ACCGCTGGTC
201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
251 ACATGGTGTG CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCG
351 GACGATAGAC AATACGTTC ACGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
15 451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGCGCG
551 CGACGCTGAC CTTTCATGACG CTTTTCGTGT GGGGGCTGTA CCGCTTCGTG
601 CCAAACCGCT TCGTTCCTCG GCGGCAGGCG TTTGTGCGGG CTTTGGCAAC
651 AGCGTTTGTG CTGGAAACCG CGCGCTCCCT CTTCACTTGG TATATGGGCA
20 701 ATTTTCGACG CTACCGCTCG ATTTACGGCG CGTTTGCCCG CGTGCCGTTT
751 TTTCTGTGTG GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CAAAGCCTTG CCTGTTCAGG AGTTCAGACG
25 951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAACCT TTCAAGCTCT TCGTTTACCG
1101 TCCGTTGCCT GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGCAGAG GTTTGACGCT
30 1201 CAGGCGAAAA AACGGCAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLQRLQGL ADNKICAF AW FVVRRFDEER VPQAAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLLV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
35 151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFMT LLLWGLYRFV
201 PNRFVPARQA FVGALATAFC LETARSLFTW YMCNFDGYRS IYGAFAAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLRHGYIY SGRQGWVLKT
351 GADSIELNEL FKLFVYRPLP VERDHNQAV DAVMTFCLQT LNMTLAEFDA
40 401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N.*

*meningitidis*:

```

45      10      20      30      40      50      60
orf144.pep MTFLQRLQGLADNKICAF AW FVVRRFDEERVPQXAASMTFTLLALVPVLTVMVAVASIF
|||||
orf144a    MTFLQRLQGLADNKICAF AW FVVRRFDEERVPQXAASMTFTLLALVPVLTVMVAVASIF
      10      20      30      40      50      60

50      70      80      90     100     110     120
orf144.pep PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLLVVTSMLLIRTID
|||||
orf144a    PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLLVVTSXMLIRTID
      70      80      90     100     110     120

60      130
orf144.pep NTFNRIWRVXXQRPWM
|||||
orf144a    NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

```

130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
5  51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTCG ACCGNTGGTC
201 GGATTTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
251 ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCG
10  351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATN GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGCGCG
15  551 CGACGCTGAN CTTCATGACG CTTTGTGCTG GGGGGCTGTA CCGCTNCGTG
601 CGAAACCGCT TCGTCCCGC GCGGCANGCG TTTGTGCGGG CTTTGGCAAC
651 AGCGTTCTGT CTGGAACCG CGCGTTCCTT CTTTACTTGG TATATGGGCA
701 ATTTCCGACG CTACCGCTCG ATTTACGGNG CGTTTGCCTG CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
20  801 GTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGNGCT
851 TCGACTCGCG CGGACGTTT GACGACGTGT TGAATACCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CNAAGCCTTG CCTGTTCAGG AGTTCAGACG
951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAACTC TTCAAGCTCT TCGTTTACCG
25  1101 TCCGTTGCCT GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGATGCCGTG TTGCGAGCT TGAACATGA CGCTGGCAGA GTTTGACGCT
1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

1  MTFLLRLQGL ADNKICAFAW FVRRFDEER VPQAAASMTF TLLALVPVL
30  51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSXMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
151 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV
201 PNRFPARXA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAPVF
25  251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RFXFDSRGRF DDVLKILLLL
301 DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLARHGYYI SGRQGWVLKT
35  351 GADSIENLNL FKL FVYRPLP VERDHVNQAV DAVMMPCLOT LNMTLAEFDA
401 QAKKQQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

```

40  orf144a.pep  MTFLLRLQGLADNKKICAFAWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF
    orf144-1    MTFLLRLQGLADNKKICAFAWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF

    orf144a.pep  PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSXMLIRTID
45  orf144-1      PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSXMLIRTID

    orf144a.pep  NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
50  orf144-1      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

    orf144a.pep  RTAATLXFMTLLLWGLYRXVPNRFVPARXAFVGALATAFCLETARSLFTWYMGNFDGYRS
    orf144-1      RTAATLXFMTLLLWGLYRXVPNRFVPARXAFVGALATAFCLETARSLFTWYMGNFDGYRS

55  orf144a.pep  IYGAFAPVFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFSRGRFDDVLKILLLL
    orf144-1      IYGAFAPVFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFSRGRFDDVLKILLLL

    orf144a.pep  DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLKTGADSIENLNL
60  orf144-1      DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLKTGADSIENLNL

    orf144a.pep  FKL FVYRPLPVERDHVNQAVDAVMMPCLOTLNMTLAEFDAQAKKQQQS 408
65  orf144-1      FKL FVYRPLPVERDHVNQAVDAVMTPCLOTLNMTLAEFDAQAKKRQ 406

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFWVVRFFDEERVQPXAASMTFTTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFWVIRRFSEERVQAAASMTFTTLLALVPVLTVMVAVASIF	60
10	orf144.pep	PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSFVSFVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144.pep	NTFNRIWRVXXQRPWM	136
15	orf144ng	NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

	1	MTFLQCWQGS	ADNKICAFW	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIVP	QGADMVFDYI	DAFRDQANRL
20	101	TAIGSVMLVV	TSMLLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPAQQA	FVGALITAF	LETARFLFTW	YMGNFDGYRS	IYGAFAAVEF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT
25	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMTPCLOT	LNMTLAEFDA
	401	QAKKQQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
	51	ATTGTCATGG	TCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
30	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG
	151	ACCGTAATGG	TCGCGGTGCG	TTCGATTTTC	CCCGTGTTCG	ACCGCTGGTC
	201	GATTTCGTT	GTCTCCTTCG	TCAACCAAAC	CATTGTGCGC	CAGGGCGCGG
	251	ATATGGTGTT	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
	301	ACCGCCATCG	GCAGCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTTCG
35	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CCTGGATGAT	GCAGTTCCTC	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCCT
	451	TTGTCTTTGG	GTGTGGGCAT	TTCCTTTATG	GTCGGGTCGG	TTCAAGACTC
	501	CCTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACGCGTTG	AAGACGGCGG
	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTTCGTG
40	601	CCCAACCGCT	TCGTGCCCGC	CCGGCAGGCG	TTTGTGCGAG	CTTTGATTAC
	651	GGCATTCTGC	CTGGAGACGG	CACGTTTCCT	GTTACCTGG	TATATGGGCA
	701	ATTTGACGCG	CTACCGCTCG	ATTTACGGCG	CATTTGCCGC	CGTGCCGTTT
	751	TTCCTGCTGT	GGTTAAACCT	GCTGTGGACG	CTGGTCTTGG	GCGGGGCGGT
45	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTTC	CGCAGGGGAT
	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTTCTG
	901	GATGCGGCGC	AAAAAGAAGG	CCGAACCTTG	TCCGTTTCAGG	AGTTTCAGACG
	951	GCATATCAAT	ATGGGTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAGCGAACTC	TTCAAGCTCT	TCGTGTACCG
50	1101	CCCGTTGCct	gtggaAAGGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
	1151	TGAcgcctg	TTTGACAGCT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGgcgAAAA	AACAGCAGCA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

55	1	MTFLQRWQGL	ADNKICAFW	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLVV	TSMLLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPAQQA	FVGALITAF	LETARFLFTW	YMGNFDGYRS	IYGAFAAVEF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
60	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT

351 GADSIELSEL FKL FVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA  
401 QAKKQQQS\*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5      orf144ng-1.pep MTFLQRWQGLADNKICAFWVIRRFSEERVQAAASMTFTTLLALVPVLTVMVAVASIF
      orf144-1      MTFLQRLQGLADNKICAFWVVRVFDEERVQAAASMTFTTLLALVPVLTVMVAVASIF

      orf144ng-1.pep PVFDRWSDSFVSFVNQTIIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
10     orf144-1      PVFDRWSDSFVSFVNQTIIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID

      orf144ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
      orf144-1      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL

15     orf144ng-1.pep KTAARLAFMTLLLWGLYRFVFNRFVPARQAFVGALITAFCLTARFLFTWYMGNFDGYRS
      orf144-1      RTAATLTFTLLLWGLYRFVFNRFVPARQAFVGALATAFCLTARSLFTWYMGNFDGYRS

20     orf144ng-1.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
      orf144-1      IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL

25     orf144ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGWVLKTGADSIELSEL
      orf144-1      DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIELNEL

      orf144ng-1.pep FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKQQQS
30     orf144-1      FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKRQ

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

1      ..AGACACGCCG GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
51     AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
101    GCACCGATAT GCGTCAGGAA ATTTCGGCCC TCGTCATCCT GCTGCAACGC
40     151    ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAG
      201    CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1      ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51     TRRKWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

1      ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
51     CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCCGG CTCCGGCGGG
101    CCGTCTGTGT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
50     151    GAGTGGATAG GGATGACCGT CTTCGTCGTC CTCGGCATGC TCCAGTTTCA
      201    AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
      251    GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCAC
      301    GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
      351    CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
      401    CGATGTGTAT GCTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
55     451    CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC

```

-352-

5 501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGCATGA CCCGCGAAGC CCTCGAGGAG AACATGGCGA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCATCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC  
 751 CGTAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA  
 10 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 1051 ACCCGCCGCA AATGGCTGGA TGCCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

15 1 MNTSQRNRLV SRWLSNYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG  
 51 EWIGMTVFVV LGMLQFQGA IYKKAVERMLG TVIGLGAGLG VLWLNQHYFH  
 101 GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCMLIG DNGSEWLD SG  
 151 LMRAMNVLIG AATAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG  
 201 RRMTRERLEE NMAKMRQINA RMVKSRSILA ATSGESRISP AMMEAMQHAH  
 20 RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING  
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR  
 351 TRRKWLDAHE RQHLRQSLE TREHG\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

					10	20	30
	orf146.pep				RHARRIRIDTAINPELEALAEHLHYQWQGF		
30	orf146a	KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF					
		280	290	300	310	320	330
		40	50	60	70		
35	orf146.pep	LWLSTDMRQEISALVILLQTRRKWLDAHERQHLRQSLETREHGX					
	orf146a	LWLSTNMRQEISALVILLQTRRKWLDAHERQHLRQSLETREHSX					
		340	350	360	370		

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40 1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCTTA  
 51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCGGC CTCGGCGGGG  
 101 CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC  
 151 GAGTGGATAG GGATGACCGT CTTCTGTCGTC CTCGGCATGC TCCAGTTTCA  
 201 AGGGGCGATT TACTCCAAGG CCGTGGAACG TATGCTCGGC ACGGTCATCG  
 251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC  
 45 301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG  
 351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCGGGGCTGA  
 401 CGATGTGCAT GCTCATCGGC GACAACGGCA GCGAATGGTT CGACAGCGGC  
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC  
 50 501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGCATGA CCCGCGAAGC CCTCGAAGAG AACATGGCGA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCACCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC  
 751 CGTAAATTTG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 55 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 60 1051 ACCCGCCGCA AATGGCTGGA TGCCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCGCTTGAA ACACGGGAAC ACAGTTGA

-353-

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

      1  MNTSQNRNLV SRWLSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
     51  EWIGMTVEFV LGMLQFQGA IYKAVRMLG TVIGLGAGLG VLWLNQHYFH
    101  GNLLFYLTVG TASALAGWAA VGKNGYVPML AGLTMCMLIG DNGSEWFDSD
5    151  LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLTDC SKMIAEISNG
     201  RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
     251  RKIVNTTELL LTAAKLQSP KLNQSEIRLL DRHFTLLQTD LQQTVALING
     301  RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
     351  TRRKWLDAHE RQHLRQSLE TREHS*

```

10 ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

```

      orf146a.pep  MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVEFV
      orf146-1     MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVEFV
15    orf146a.pep  LGMLQFQGA IYKAVRMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
      orf146-1     LGMLQFQGA IYKAVRMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
20    orf146a.pep  VGKNGYVPMLAGLTMCLIGDNGSEWFDSDGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
      orf146-1     VGKNGYVPMLAGLTMCLIGDNGSEWFDSDGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
25    orf146a.pep  FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
      orf146-1     FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
30    orf146a.pep  AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNQSEIRLLDRHFTLLQTDLQQTVALING
      orf146-1     AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNQSEIRLLDRHFTLLQTDLQQTVALING
35    orf146a.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTRRKWLDAHE
      orf146-1     RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTRRKWLDAHE
40    orf146a.pep  RQHLRQSLE TREHSX
      orf146-1     RQHLRQSLE TREHGX

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

```

      orf146.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF 30
      orf146ng   KLNGSEIRLLDRHFTLLQTDLQQTAAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF 364
45    orf146.pep  LWLSTDMRQEISALVILLQRTRRKWLDAHERQHLRQSLE TREHG 75
      orf146ng   LWLSTNMRQEISALVIPLQRTRRKWLDAHERQHLRQSLE TREHG 409

```

50 An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

      1  MSGVRFPSA PIPSTDPPSG SLCFFTFPLQ TASDMNSSQR KRLSGRWLNS
     51  YERYRHRRLI HAVRLGGTVL FATALARLLH LQHGEWIGMT VFVVLGMLQF
    101  QGA IYSSNAVE RMLGTVIGLG AGLGVWLNQ HYFHGNLLFY LTIGTASALA
55   151  GWA AVGKNGY VPMLAGL TMC MLIGDNGSEW LDSGLMRAMN VLIGAAIAIA
     201  AAKLLPLKST LMWRFMLADN LADCSKMIAE ISNGRRMTRE RLEQNMVKMR
     251  QINARMVKSR SHLAATSGES RISPSMMEAM QHAHRKIVNT TELLTAAK
     301  LQSPKLNQSE IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
     351  EALAEHLHYQ WQGF LWLSTN MRQEISALVI PLQRTRRKWL DAHERQHLRQ
     401  SLE TREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

```

      1 ATGAACCTCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA
    51 CGAACGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCgga
  101 ccgtCCTGTT CGCCACCGCA CTCGCCCGgc tACTCCACCT CCAacacggc
  151 gAATGGATAG GGAAtgaCCGT CTTCTGTCGTC CTCGGCATGC TCCAGTTCCA
    201 AGGCgcgatt tActccaacg cgggtgGAacg taTGctcggt acggtcatcg
    251 ggctgGGCGC GGGTTTGGgc gTTTTATGGC TGAACCAGCA TTAtttccac
    301 ggcaacCTcc tcttctacct gaccatcggc acggcaagcg cactggccgg
    351 ctGGGCGCGC GTCGGCAAAA acggctacgt ccctatgctg GCGGGGctgA
  401 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC
  451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCGC
  501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
  551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
  601 AGCGGTATGA CGCGCGAACG TTTGGAGCAG AATATGGTCA AAATGCGCCA
  651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCACCTCGCC GCCACATCGG
  701 GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCCC
  751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
  801 GCAATCTCCC AAACCAACG GCAGCGAAAT CCGGCTGCTC GACCGCCACT
  851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC
  901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAATGGA
  951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

```

  1 MNSSQKRRLS GRWLNSEYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG
  51 EWIGMTVFVV LGMLQFQGA IYNAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTIG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLD SG
 151 LMRAMNVLIQ AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
 201 RRMTRERLEQ NMVKMRQINA RMVKSRSHLA ATSGESRISP SMMEAMQHAH
 251 RKIVNTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTAAALING
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRKWLDAHE RQHLRQSLLE TREHG*

```

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

```

 35 orf146-1.pep MNTSQRNRLVSRWLNSEYERYRHRRLIHAVRLGGAVLFATASARLLHLQHG
    orf146ng-1 MNSSQKRRLSGRWLNSEYERYRHRRLIHAVRLGGTVLFATALARLLHLQHG
    orf146-1.pep LGMLQFQGA IYNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
    orf146ng-1 LGMLQFQGA IYNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
 40 orf146-1.pep VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIQAAIAIAAAKL LPLKSTLMWR
    orf146ng-1 VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIQAAIAIAAAKL LPLKSTLMWR
 45 orf146-1.pep FMLADNLADC SKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
    orf146ng-1 FMLADNLADC SKMIAEISNGRRMTRERLEQNMVKMRQINARMVKSRSHLAATSGESRISP
 50 orf146-1.pep AMMEAMQHAHRKIVNTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
    orf146ng-1 SMMEAMQHAHRKIVNTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTAAALING
 55 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
    orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
    orf146-1.pep RQHLRQSLLE TREHG
    orf146ng-1 RQHLRQSLLE TREHG
 60

```

Furthermore, ORF146ng-1 shows homology with a hypothetical *E.coli* protein:

```

 65 sp|P33011|YEEA_ECOLI_HYPOTHETICAL_40.0_KD_PROTEIN_IN_COBU-SBMC_INTERGENIC_REGION
    >gi|1736674|gnl|PID|d1016553 (D90838) ORF_ID:o348#20; similar to [SwissProt
    Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)
    ORF_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

```



-355-

>gi|1788318 (AE000292) f352; 100% identical to fragment YEEA\_ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352  
 Score = 109 bits (271), Expect = 2e-23  
 Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

5 Query: 20 YRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAIYSNAVERML 79  
 YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+

Sbjct: 15 YRHYRIVHGTRVALAFLLTFLIIRLFTIPESTWPLVTMVVIMGPISFWGNVVPRAFERIG 74

10 Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGNKYVPMLAGLTMCMLI 139  
 GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++

Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAMFLCGWLALGKKPYQGLLIGVTIAIVV 131

15 Query: 140 GDNGSEWLDSGLMRAMNVLIGXXXXXXXXXKLLPLKSTLMWRFMLADNLADCSKMI AEISN 199  
 G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +

Sbjct: 132 GSPTGE-IDTALWRS GDVILGSL LAMLEFTGIW PQRAFIHWRIQLAKSLTEYNRVYQSAFS 190

20 Query: 200 GRRMTRERLEQNVMKMRQINARMVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259  
 + R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V

Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKETRI PKSIYEGIQ TINRNLCMLLEL 247

25 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXAALINGRHARRIRIDTAINPEL 316  
 + LN ++R D AL G +N +

Sbjct: 248 QINAYWATRPSHFVLLNAQKLR--DTQHMMQ QILLSLVHALYEGNPQPVFANTEKLNDV 305

Query: 317 EALAEHL--HYQWQ-----GFLWLSTNMQR EISALVILLQRTTRK 354  
 E L + L H+ + G++WL+ ++ L L+ R RK

Sbjct: 306 EELRQLLNHHDLKVVETPIYGYVWLNMETAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the  
 30 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

35 1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA  
 51 GGGCAAATC GTCAGTGTGC GCGAACACAA CGAACGGCAG ATGGCGGACA  
 101 AGATTGTCCG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTTCCGAT  
 151 GCGGGTACGC CGGCGGTGTG CGACCGGGG GCGAACTCG CCCGCCGCGT  
 201 GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA  
 251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATT TTATTTC AAC  
 40 301 GGTTTGTAC CGCCGAAATC GGGAGAAGCG AGGAACTGT TTGCCAAATG  
 351 GGTGCGGGCG GCGTTTCTTA TCGTCATGTT TGAAACGCG CACCGCATCG  
 401 GTGCAGCGCT TGCCGATATG GCGGAACTGT TCCCCGAACG CCGATTAATG  
 451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT  
 501 TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG  
 45 551 AGATGGTGTT GGTGCTTTAT CCGCGCGAGG ATGAAAAACA CGAAGGCTTG  
 601 TCCGAGTCCG CGCAAAACAT CATGAAAATC CTCACAGCCG AGCTGCCGAC  
 651 CAAACAGGCG GCGGAGCTTG CTGCCAAAAT CACGGCGGAG GGAAAGAAAG  
 701 CTTTGTACGA T..

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

50 1 ..AEDTRVTAQL LSAYGIQGL VSVREHNERQ MADKIVGYLS DGMVVAQVSD  
 51 AGTPAVCDPG AKLARRVREA GFKVVPVGA XAVMAALSVA GVEGSDFFYN  
 101 GFVPPKSGER RKLFAKWVRA AFPIVMFETP HRIGAALADM AELFPERRIM  
 151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHEGL  
 201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKKALYD..

55 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTCAGA AACATTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC  
 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
 151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATT CAGG CAAACTCGT

-356-

5  
10  
15

```

201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGCGCT
251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTCAACGG TTTTGTACCG
451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TCGGGCGGCC
501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GCGCGCGCAA
601 ATTACGAAAA CGTTTGAACG GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGCGCAG ATGGTGTGTTG
701 TGCTTTATCC GCGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751 CAAAACATCA TGAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801 GGAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
851 TGGCTCTGTC TTGAAAAAC AAATAG

```

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

20

```

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQKLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP
151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRLMLARE
201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E.coli* (accession number U18997)

ORF147 and *E.coli* ORF286 protein show 36% aa identity in 237aa overlap:

25

```

Orf147: 1 AEDTRVTAQLLSAYGIQKLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
Orf286: 43 AEDTRHTGLLLQHFGINARLFA LHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPG 102

Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLF AKWVRA 120
L R RE F + GF+P KS RR
Orf286: 103 YHLVRTCREAGIRVVPLPGPCAAITALS AAGLPSDRFCYEGFLPAKSKGRRDALKAIEAE 162

Orf147: 121 AFPIVMFETPHRIG AALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179
++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
Orf286: 163 PRTLIFYESTHRL LDSLEDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDEN 222

Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELAAKITGEGKKALY 236
+ +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALLQAE LPLKAAALAEIHGVKKNALY 278

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

45

```

          10      20      30
orf147.pep      AEDTRVTAQLLSAYGIQKLVSVREHNERQ
|||
orf75a      TLYVVATPIGNLADITLRLALAVLQKADIICAEDTRVTAQLLSAYGIQKLVSVREHNERQ
          20      30      40      50      60      70

          40      50      60      70      80      90
orf147.pep      MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXVMAALSVA
|||
orf75a      MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASVMAALSVA
          80      90      100      110      120      130

          100      110      120      130      140      150
orf147.pep      GVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFPIVMFETPHRIG AALADMAELFPERRLM
|||
orf75a      GVAGSDFYFNGFVPPKSGERRKLF AKWVRVAFVPMFETPHRIGATLADMAELFPERRLM
          140      150      160      170      180      190

          160      170      180      190      200      210
orf147.pep      LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI

```

```

|||||
orf75a      LAREITKTFETFLSCTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI
          200      210      220      230      240      250

          220      230
orf147.pep  LTAELOPTKQAAELAAKITGEGKKALYD
          |||||
orf75a      LTAELOPTKQAAELAAKITGEGKKALYDLALSWNKX
          260      270      280      290

```

Homology with a predicted ORF from *N.gonorrhoeae*

15	orf147.pep	AEDTRVTAQLLSAYGIQKKLVSVREHNERQ	30
	orf147ng	TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGRLVSVREHNERQ	85
20	orf147.pep	MADKIVGYLSDGMVVAQVSDAGT PAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA	90
	orf147ng	MADKIVIGFLSDGLVVAQVSDAGT PAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA	145
25	orf147.pep	GVEGSDIFYFNGFVPPKSGERRKLF AKWVRAAFPVVMFETPHRIGAAALDMAELFPERRLM	150
	orf147ng	GVAESDIFYFNGFVPPKSGERRKLF AKWVRAAFPVVMFETPHRIGATLADMAELFPERRLM	205
30	orf147.pep	LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI	210
	orf147ng	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNAMKI	265
30	orf147.pep	LTAE LPTKQAAELAAKITGEGKKALYD 237	
	orf147ng	LAELPTKQAAELAAKITGEGKKALYDLALSWKNK 300	

35	1	MSVFQTAFFM	FQKHLQKASD	SVVGGTLYVV	ATPIGNLADI	TLRALAVLQK
	51	ADIICAEDTR	VTAQLLSAYG	IQGRLVSVRE	HNERQMAKV	IGFLSDGLVV
	101	AQVSDAGTFA	VCDPGAKLAR	RVREAGFKVV	PFVVASAVMA	<u>ALS</u> VAGVAES
	151	DFYFNGFVPP	KSGERRKLFA	KWVRAAFVTV	MEETPHRIGA	TLADMAELFP
40	201	ERRLMLAREI	TKTFETFLSC	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE
	251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWNKN
	301	*				

45	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCCTGC
	101	CGCCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTTTGTG	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GCCATTTCAGG	GCAGGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
50	251	TCCTTTTACA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGT	GGGTACGCCG
	301	GCGGTGTGCG	ATCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGCGCGAAG	CGCGGTAATG	GCGCGCTTGA
	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTTGTT	GCCAAATGGG	TGCGGGCGCG
55	501	ATTTCTCTGT	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
	551	CCGATATGGC	GGAATTGTTT	CCCGAACGCC	GTCTGATGCT	GGCGCGCGAA
	601	ATCACGAAAA	CGTTTGAAAC	GTTCTTAAAG	GGCACGGTTG	GGGAAATTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGGG
	701	TGCTTTATCC	GCGCGAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
60	751	CAAAATCTGA	TGAAAATCCT	TGCGGCGGAG	CTGCCGACCA	AGCAGGCGGC
	801	GGAGCTTGCC	GCCAAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	GTGGAAAAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

      1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
    51 RVTAQLLSAY GIQGRVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP
   101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
    151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF PERRMLLARE
    201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
    251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWNK K*

```

ORF147ng shows homology to a hypothetical *E. coli* protein:

```

10  sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
    (F286)
    >gi|606086 (U18997) ORF_f286 [Escherichia coli]
    >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
    [Escherichia coli] Length = 286
    Score = 218 bits (550), Expect = 3e-56
    Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

    Query: 4 KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
            K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
    20  Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVQLAVDLIAEDTRHTGLLLQHFGIN 59

    Query: 64 GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
            RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
    25  Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

    Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFQKWRVRAAFVVMFETPHRIGATL 183
            G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
    30  Sbjct: 120 PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRDALKAEAEPRTLIFYESTHRLDLSL 179

    Query: 184 ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
            D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
    35  Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGE LLAWVKEDENRRKGEMVLIV-EGHKAQ 238

    Query: 243 HEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLAL 286
            E L A + +L AELP K+AA LAA+I G K ALY AL
    40  Sbjct: 239 EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 77

40 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 647>

```

      1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
    51 AACCGGTGCG ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCGT
   101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
    151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
    201 GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
    251 CAATGACAAA AGCCCCGATG ATTGATTTT CTGTGGTGTC GCGTAACGGC
    301 GTGGCGGCAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG
    351 GCGGCTATAA CAACGTTGAT TTTGGTGCGG AAGGAak.AA tATCCC.GAT
    401 CAACAwCGwW TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
    451 GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGATA
    501 AATwTGTCAC AGATGCAGAA CCTGTTGAAA TGACCAAGTTA TATGGATGGG
    551 CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTT GTATTGGGGC
    601 AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACCAGCGAAA
    651 GTTCATATCA TATTGCAAGT .....
    701 ..... GGCTC ACCAATGTTT ATCTATGATG CCAAAAAGCA
    751 AAAGTGTTTA ATTAATGGGG TATTGCAAAAC GGGCAACCCC TATATAGGAA
    801 AAAGCAATGG CTTCCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC
    851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACCACGTC AAAATGGGAA
    901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC

```

-359-

951 ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG  
 1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC  
 1051 AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAATATTTT  
 1101 CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT  
 1151 CAAGGTGCTG GAGGATTATA TTTCGAAGGA GATTTTACGG TCTCGCCTGA  
 1201 AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA  
 1251 CCGTTACTTG GAAAGTAAAC GGCGTGGCAA ACGACCGCCT GTCCAAAATC  
 1301 GGCAAAGGCA CGCTG.....  
 110 2101 .....  
 2151 TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC  
 2201 GATCAGCTC ATTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT  
 2251 TAGTGCAAAT GGCGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA  
 15 2301 ACGGCAACCK TAAGCCTCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC  
 2351 ACATTAAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG  
 2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCGGC AACGCTAAG  
 2451 CAAACGTAAG CCATTCCGCA CTCAACGGTA ATGTCTCCCT AGCCGATAAG  
 2501 GCAGTATTCC ATTTTGAAAG CAGCCGCTTT ACCGGACAAA TCAGCGGCGG  
 2551 CAAGGATACG GCATTACACT TAAAAGACAG CGAATGGACG CTGCCGTCag  
 2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT  
 2651 TCCGCCTATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGA  
 2701 TCGCGCGCGC CGCCGTTCGC GCCGTTCGCG CCGTTCCTTA TTATmCGTTA  
 2751 CACCGCCAAC TTCGGTAGAA TCCCGTTTCA ACACGCTGAC GGTAACCGGC  
 2801 AAATTGAACG GTCAGGGAAC ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA  
 25 2851 CCGCAGCGAC AAATTGAAGC TGGCGGAAAG TTCCGAAGGC ACTTACACCT  
 2901 TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG  
 2951 GTAGTGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTAC  
 3001 CCTGCAAAAC GAACACGTCG ATGCAGGCGC GTGG.....  
 30 3551 .....  
 3601 CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAA CACTACCGTT  
 3651 CGCAAGATTT CCGCGCCTAC CGCCAACAAA CCGACCTGCG CCAAATCGGT  
 3701 ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC GGCATCCTGT TTTCGCACAA  
 3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CCGCAACTCG GCACGGCTTG  
 35 3801 CCCACGGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC  
 3851 ATCAGnCGCG GCGCGGGTT TAGCAGCGG CAGCCTTtCa GACGCGATCG  
 3901 GAGsmAAAwT CCGCCGCCGC GTGctGCATT ACGGCATTCA GGCACGAtAC  
 3951 CGCGCCGgtt tCgGcGgAtt CGGCATCGAA CCGCACATCG GCGCAACGCG  
 40 4001 ctATTTCGTC CAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA  
 4051 CCCCCGGCCT TGCAATCAAC CGcTACCGCG CGGGCATTAA GGCAGATTAT  
 4101 TCATTCAAAC CGGCGCAACA CATTTCCATC ACGCCTTATT TGAGCCTGTC  
 4151 CTATACCGAT GCCGCTTCGG GCAAAGTCCG AACACGCGTC AATACCGCCG  
 4201 TATTGGCTCA GGATTTCGGC AAAACCCGCA GTGCGGAATG GGgCGTAAAC  
 4251 GCCGAAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCCAAAGG  
 45 4301 CCGCAACTG GAAGCGCAAC ACAGCGCGGG CATCAAATTA GGCTACCGCT  
 4351 GGTAA...

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

1 MKTTDKRTTE THRKAAPTGR IRFXAAYLAI CLSFGILPOA WAGHTYFGIN  
 50 51 YQYYRDFEEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRG  
 101 VAALVGQYI VSAHNGGYN NVDFGAEGXN IXDQXRXYK IVKRNNYKAG  
 151 TKGHPYGGDY HMPRLHKXVT DAEPVEMTSY MDGRKYIDQN NYPDRVRIGA  
 201 GRQYWRSDDED EPNNRESSYH IAS.....GS PMFIYDAQKQ  
 251 KWLINGVLQT GNPYIGKSNG FQLVRKDWFY DEIFAGDTHS VFYEPRONGK  
 301 YSFNDNNGT GKINAKHEHN SLPNRLKTRT VQLEFNVSLE TAREPVYHAA  
 55 351 GGVNSYRPRL NNGENISFID EGKGELILTS NINQAGGLY FQGDFTVSPE  
 401 NNETWQAGAV HISEDSTVTW KVNGVANDRL SKIGKGTLL.....  
 701 .....DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL  
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS  
 801 DHAVQNGSLT LSGNANANVS HSALNGNVS ADKAVFHFES SRFTGQISGG  
 851 KDTALHLKDS EWTLP SGXEL GNLNLDNATI TLNSAYRHDA AGAQTGSATD  
 901 APRRRSRRSR RSLLVTPPT SVESRENTLT VNGKLNGQGT FRFMSSELFY  
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVEGKDN KPLSENLFY  
 1001 LQNEHVDAGA W.....  
 65 1151 .....LDRVFAEDR  
 1201 RNAVWTSRIR DTKHYRSQDF RAYRQOTDLR QIGMOKNLGS GRVGILFSHN  
 1251 RTENTFDDGI GNSARLAHGA VFGQYQIDRF YIGISAGAGF SSGSLSDGIG  
 1301 XKXRRRVLHY GIQARYRAGF GFGIEPHIG ATRYFVQKAD YRYENVNIA  
 70 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDAAAS KVRTRVNTAV

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW  
1451 \*

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

```

5      1 ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
      51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCTG
     101 TCGGCATCTCT TCCCCAAGCC TGGGCGGGAC ACACCTATTT CGGCATCAAC
     151 TACCAATACT ATCGCGACTT TGCCGAAAAA AAAGGCAAGT TTGCAGTCGG
     201 GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAAT
     251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
     301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
     351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC
     401 ATCGTTTTAC TTATAAAATT GTGAAACGGA ATAATTATAA AGCAGGGACT
     451 AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCCGCTT TGCATAAATT
     501 TGTCACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
     551 AATATATCGA TCAAAATAAT TACCCTGACC GTGTTCTGAT TGGGGCAGGC
     601 AGGCAATATT GGCGATCTGA TGAAGATGAG CCCAATAACC GCGAAAGTTC
     651 ATATCATATT GCAAGTGCCT ATTCTTGGCT CGTTGGTGGC AATACCTTTG
     701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG TGAAAAAATT
     751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
     801 TGGCTCACCA ATGTTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
     851 ATGGGGTATT GCAAAAGGGC AACCCCTATA TAGGAAAAAG CAATGGCTTC
     901 CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
     951 CCATTAGTA TTCTACGAAC CACGTCAAAA TGGGAATAAC TCTTTTAAAC
    1001 ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT
    1051 CTGCCATAAT GATTAATAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
    1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA
    1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTCT TATTGACGAA
    1201 GGAAAAGGCG AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG
    1251 ATTATATTTT CAAGGAGATT TTACGGTCTC GCCTGAAAAA AACGAACTT
    1301 GGCAAGGCGC GGGCTTTCAT ATCAGTGAAG ACAGTACCGT TACTTGGAAA
    1351 GTAAACGGCG TGGCAAAACGA CCGCTGTCTC AAAATCGGCA AAGGCACGCT
    1401 GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG
    1451 GTACAGTCAT TTTGGATCAG CAGGCAGACG ATAAAGGCAA AAAACAAGCC
    1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGTACGGTGC AACTGAATGC
    1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTTCGGCTT CGCGGCGGAC
    1601 GTTTGGATTT AAACGGGCAT TCGCTTTCGT TCCACCGTAT TCAAAATACC
    1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
    1701 TACCAATTACA GGCAATAAAG ATATTGCTAC AACCAGCAAT AACACAGAGT
    1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
    1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACCAGC CCGCCGCAGA
    1851 AGACCGCACC CTGCTGCTTT CCGGCGGAAC AAATTTAAAC GGCAACATCA
    1901 CGCAAAACAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC
    1951 TACAATCATT TAAACGACCA TTGGTTCGCA AAAGAGGGCA TTCCTCGCGG
    2001 GGAAATCGTG TGGGACAACG ACTGGATCAA CCGCACATTT AAAGCGGAAA
    2051 ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCCCGCAA TGTTGCCAA
    2101 GTGAAAGGCG ATTTGGCATT GAGCAATCAC GCCCAAGCAG TTTTGTGTTG
    2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
    2201 TGACAAATTG TGTGAAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
    2251 TTGACTAAGA CCGACATCAG CGGCAATGTC GATCTTGCCG ATCAGCTCA
    2301 TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAAATG
    2351 GCGATACACG TTATACAGTC AGCCACAACG CCACCAAAAA CGGCAACCTT
    2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAACCGG
    2451 CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG
    2501 TACAAAACGG CAGTCTGACG CTTTCGGGCA ACGCTAAGGC AAACGTAAGC
    2551 CATTCCGCAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
    2601 TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGCGCGC AAGGATACGG
    2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCAGG CACGGAATTA
    2701 GGCAATTTAA ACCTTGACAA CGCCACCATT ACACCTCAAT CCGCCTATCG
    2751 CCACGATGCG GCAGGGGCGC AAACCGGCAG TCGCAGAGAT GCGCCGCGCC
    2801 GCCGTTGCGC CCGTTGCGCG CGTTCCTTAT TATCCGTTAC ACCGCCAACT
    2851 TCGGTAGAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG
    2901 TCAGGGAACA TTCCGCTTTA TGTGGAACCT CTTTCGGCTAC CGCAGCGACA
    2951 AATTGAAGCT GCGGAAAGT TCCGAAGGCA CTTACACCTT GCGGTCACAC
    3001 AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGGGAGG
    3051 AAAAGACAAC AAACCGCTGT CCGAAAACCT TAATTTACC CTGCAAAACG
    3101 AACACGTCGA TGCCGCGCGC TGGCGTTACC AACTCATCCG CAAAGACGGC
    3151 GAGTTCGGCC TGCATAATCC GGTCAAAGAA CAAGAGCTTT CCGACAAACT
    3201 CGGCAAGGCA GAAGCCAAAA AACAGGCGGA AAAAGACAAC GCGCAAGGCC
    3251 TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAAA GACAGAAAGC
    3301 GTTGCCGAAC CGGCCCGGCA GGCAGGCGGG GAAAATGTGC GCATTATGCA

```

-361-

5  
10  
15  
20

```

3351 GGC GGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCCTTGG
3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTTCCCCCGC
3451 GCGCGCCGCG CCCGCCGGGA TTGCGCGCAA CTGCAACCCC AACCGCAGCC
3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGTTTGTAGTG
3551 AATTTTCCGC CACGCTCAAC AGCGTTTTCG CCGTACAGGA CGAATTAGAC
3601 CGCGTATTTG CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
3651 GGACACCAAA CACTACCGTT CGCAAGATTT CCGCGCCTAC CGCCAACAAA
3701 CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC
3751 GGCATCCTGT TTTCGCACAA CCGGACCGAA AACACCTTCG ACGACGCAT
3801 CGGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTTCGG CAATACGGCA
3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGGC
3901 AGCCTTTCAG ACGGCATCGG AGGCAAAATC CGCCGCGCG TGCTGCATTA
3951 CGGCATTCAG GCACGATACC GCGCCGTTT CCGCGGATTG GGCATCGAAC
4001 CGCACATCGG CGCAACGCGC TATTTCTGTC AAAAAGCGGA TTACCGCTAC
4051 GAAAACGTCA ATATCGCCAC CCCCGGCCCT GCATTCAACC GCTACCGCGC
4101 GGGCATTAAG GCAGATTATT CATTCAAACC GCGCAACAC ATTTCCATCA
4151 CGCCTTATTT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4201 ACACGCGTCA ATACCGCGT ATTGGCTCAG GATTTTCGGA AAACCGCGAG
4251 TGCGGAATGG GCGGTAAACG CCGAAATCAA AGGTTTCACG CTGTCCCTCC
4301 AGCCTGCCGC CGCCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
4351 ATCAAATTAG GCTACCGCTG GTAA

```

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

25  
30  
35  
40  
45  
50

```

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYYRDFEEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG
101 VAALVGDQYI VSVAHNNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
151 KGHYPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
201 RQYWRSEDEDE PNNRESSYHI ASAYSWLVG GNTFAQNGSGG GTVNLGSEKI
251 KHSPYGFLEPT GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNFG
301 QLVRKDWFYD EIFAGDTHSV FYEPRQNGKY SFNDDNNGTG KINAKHEHNS
351 LPNRLKTRTV QLFNVSLSET AREPVYHAA GVNYSRPRLN NGENISFIDE
401 KGCELILTSN INQAGGGLYF QGDFTVSPEN NETWQAGGVH ISEDSTVTWK
451 VNGVANDRLS KIGKGTLLHVQ AKGENQGSIS VGDGTVILDQ QADDKCKKQA
501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
551 DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFEKED
601 TTKTNRLNLL VYQPAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
651 YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
701 VKGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
801 SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
851 HSAALNGVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSTGTEL
901 GNLNLNDNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
951 SVESRENTLT VNGKLNGQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN
1001 NTGNEPASLE QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG
1051 EFRLHNPVKE QELSDKLGKA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
1101 VAEPARQAGG ENVGIMQAE EKKRVQADKD TALAKQREAE TRPATTAFFR
1151 ARRARDLPQ LQPQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQDELD
1201 RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQOTDLRQIG MQKNLGSGRV
1251 GILFSHNRTE NTFDDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG
1301 SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAAAGKVR
1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
1451 IKLGYRW*

```

Computer analysis of these sequences gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

60

```

      10      20      30      40      50      60
orfl1.pep MKTTDKRTE THRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGIN YQYYRDFEEN
|||||
orfl1a    MKTTDKRTE THRKAPKTGRIRFSPAYLAI CLSFGILPQAWAGHTYFGIN YQYYRDFEEN
      10      20      30      40      50      60
      70      80      90     100     110     120
orfl1.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRNGVAALVGVQYIVSVAHNNGGYN

```

5

15

20

25

30

35

40

45

50

55

60

65

70

orf1a	KGKFAVGA	KDIEVYN	KKGELV	GKSMTK	APMIDF	SVVSR	NGVAAL	VG	GDQY	IVSVA	HNGGYN
	70	80	90	100	110	120					
orf1.pep	NVDFGA	EGXNIX	DQXRXY	KIVKR	NNYKAG	TKGHPY	GGDYH	MPRLH	KXVTD	AE	PVEMTSY
orf1a	NVDFGA	EGXN-PD	QHRFSY	QIVKR	NNYKPD	NS-HPY	NGDXH	MPRLH	KFVTD	AE	PVEMTSD
	130	140	150	160	170						
orf1.pep	MDGRKY	IDQNNYP	DPRVR	IGAGR	QYWR	SDEDEF	-----	NN-----			
orf1a	MRGNTY	SDKEKY	PERVR	IGSGH	HYWRY	DDDKH	GDLSY	SGAWL	IGGNTH	MQGW	NNGVXSL
	180	190	200	210	220	230					
orf1.pep	-----	RESSYH	-----	IA-----	SGSPM	FIYDA	QKQK	WLING	VLQTG	NPYIG	KSNGFQLVRK
orf1a	SGDVR	HANDYG	PMPIA	GAGDS	SGSPM	FIYDK	TNNK	WLLNG	VLQTG	YPYSG	RENGFQLIRK
	240	250	260	270	280	290					
orf1.pep	DWFYDE	IFAGD	THSVFY	EPRQNG	KYSFN	DDNNGT	GKINAK	HEHNS	LPLN	RKLK	TRTVQLFNV
orf1a	DWFYDD	IYRGD	THTVX	FEP	SRNGH	FSFTS	NNNGT	GTVTET	NEKVS	NP-KL	KVQTVRLFDE
	300	310	320	330	340	350					
orf1.pep	SLSETA	REP	PVYHA	AGGVNS	YRPR	LNNGEN	ISFIDE	GKGEL	ILTSN	INQG	AGGLYFQGDFT
orf1a	SLNETD	KEPVY	-AAG	GVNQY	RPR	LNNGEN	LSFIDY	GNGKL	ILSN	INQG	AGGLYFEGDFT
	360	370	380	390	400	410					
orf1.pep	VSPENNE	TWQGA	GVHISE	DSTVT	WKVNG	VANDRL	SKIGK	GTL-----			
orf1a	VSPENNE	TWQGA	GVHISE	DSTVT	WKVNG	VANDRL	SKIGK	TLHVQ	AKGEN	QGSIS	VG
	420	430	440	450	460	470					
orf1.pep	-----										
orf1a	VILDQ	QADDK	GKKQAF	SEIGL	XSGR	GT	VLNAD	NQFNP	DKLYF	GFRG	RLDLNGHSLSFH
	480	490	500	510	520	530					
orf1.pep	-----										
orf1a	RIQNT	DEGAM	IXXHN	ATTTST	VTITG	NESITQ	PSGKN	INRLN	YSKEI	AYNGW	FGEKDTTK
	540	550	560	570	580	590					
orf1.pep	-----										
orf1a	TNGRLN	LVYQ	PAEDRT	XL	SSG	TNLNG	NITQT	NGKL	FFSG	RPTPH	AYNHLSGSGWSKMEG
	600	610	620	630	640	650					
orf1.pep	-----										
orf1a	IPQGE	I	VDNDW	IXRTF	KAENF	HIQGG	QAVIS	RNVAK	VEGD	XHLSN	HAQAVFGVAPHQSH
	660	670	680	690	700	710					
orf1.pep	-----										
orf1a	TICTR	SDWTG	L	TNCV	EXXIT	DDKVI	ASLTK	TDXSG	VXLX	XXXXXX	LXGXAXLXGNLSAN
	720	730	740	750	760	770					
orf1.pep	GDTRY	TVSHN	ATQNG	NXSL	VNAQ	ATFNQ	ATLNG	NTSAS	GNAS	FNLS	DHAVQNGSLTLS
	490	500	510	520	530	540					



-363-

```

|||||
orfla  GDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAQNGSLTLSD
      780      790      800      810      820      830

5      550      560      570      580      590      600
orfl.ppep NAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGNL
      |||||||
orfla     NAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSGTTELGNL
      840      850      860      870      880      890

10     610      620      630      640      650      660
orfl.ppep NLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRSLLXVTPPTSVESRFTLTVNG
      |||||||
orfla     NLDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---LLSVTPPTSVESRFTLTVNG
      900      910      920      930      940      950

15     670      680      690      700      710      720
orfl.ppep KLNGQGTFRFMSELEFGYRSDKLKLAESSEGYTLAVNNTGNEFASLEQLTVVEGKDNKPL
      |||
orfla     KLNQGTFRFMSELEFGYRSDKLKLAESSEGYTLAVNNTGNEFPVSLDQLTVVEGKDNKPL
      960      970      980      990      1000      1010

20     730      740      750
orfl.ppep SENLNFLLQNEHVDAGAW-----
      |||||||
orfla     SENLNFLLQNEHVDAGAWRYQLIRKDGELRLHNPVKEQELSDKLGKAEAKKQAEKDNAQS
      1020      1030      1040      1050      1060      1070

30     orfl.ppep -----
orfla     LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKKRVQADKDSALAKQREAE TRP
      1080      1090      1100      1110      1120      1130

35     orfl.ppep -----760-----LDR
orfla     XTTFAPPRARXARRDLPPQPQPQPQPQORDLXSRYANSGLSEFSATLNSVFAVQDELDR
      1140      1150      1160      1170      1180      1190

40     770      780      790      800      810      820
orfl.ppep VFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNR TEN
      |||||||
orfla     VFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNR TEN
      1200      1210      1220      1230      1240      1250

45     830      840      850      860      870      880
orfl.ppep TFDDGIGNSARLAHGA VFGQYIDRFYIGISAGAGFSSGSLSDGIGXKXRRRVLHYGIQA
      :
orfla     XFDDGIGNSARLAHGA VFGQYIGIRFDIGISTGAGFSSGXLSDGIGGKIRRRVLHYGIQA
      1260      1270      1280      1290      1300      1310

50     890      900      910      920      930      940
orfl.ppep RYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
      |||||||
orfla     RYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHX
      1320      1330      1340      1350      1360      1370

55     950      960      970      980      990      1000
orfl.ppep SITPYLSLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAA AKGP
      |||||||
orfla     SITPYXLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHAAA AKGP
      1380      1390      1400      1410      1420      1430

60     1010      1020
orfl.ppep QLEAQHSAGIKLGYRWX
      |||||||
orfla     QLEAQHSAGIKLGYRWX
      1440      1450

65
70 The complete length ORF1a nucleotide sequence <SEQ ID 651> is:

```

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCGT
	101	TCGGCATTCT	TCCCCAAGCT	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC
5	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTNT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAAT
	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCGTAACGGC
	301	GTGGCGGCAT	TGGTGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AACGTTGATT	TTGGTGCGBA	AGGAAGNAAT	CCCGATCAGC
	401	ACCGTTTTTC	TTACCAAATT	GTGAAAAGAA	ATAATTATAA	GCCTGACAAAT
10	451	TCACACCCCTT	ACAACGGCGA	TTANCATATG	CCGCGTTTGC	ATAAATTTGT
	501	CACAGATGCA	GAACCTGTCT	AAATGACGAG	TGACATGAGG	GGGAATACCT
	551	ATTCGGATAA	AGAAAAATAT	CCCGAGCGTG	TCCGCATCGG	CTCAGGACAC
	601	CACATATGGC	GTTATGATGA	TGACAAACAC	GGCGATTAT	CCTACTCCGG
	651	CGCATGGTTA	ATTGGCGGCA	ATACACATAT	GCAGGGTTGG	GGAAATAATG
15	701	GCGTANTTAG	TTTGAGCGGC	GATGTGCGCC	ATGCCAACGA	CTATGGCCCT
	751	ATGCCGATTG	CAGGTGCGGC	AGGCGACAGC	GGTTCGCCAA	TGTTTTATTTA
	801	TGACAAAACA	AACAATAAAT	GGCTGCTCAA	CGGAGTTTTA	CAAACCGGCT
	851	ACCCTTATTC	CGGCAGGGAA	AACGGTTTCC	AGCTGATACG	CAAAGATTGG
	901	TTCTACGATG	ACATTTACAG	AGGCGATACA	CATACCGTCT	NTTTTGAACC
20	951	GGCAGTAAC	GGACATTTTT	CCTTTACATC	CAACAACAAC	GGTACGGGTA
	1001	CGGTAACAGA	AACCAACGAA	AAGGNTNCCA	ATCCAAGGCT	TAAAGTACAG
	1051	ACAGTCCGAC	TGTTTGACGA	ATCTTTGAAT	GAAACTGATA	AAGAACCAGT
	1101	TTACGCGGCA	GGGGGTGTTA	ATCAGTACCG	TCCAAGGTTA	AACAACGGTG
25	1151	AAAACCTTTC	TTTTATCGAT	TACGGCAACG	GCAAACTCAT	CTTATCAAAAT
	1201	AACATCAACC	AAGGCGCGGG	CGGTTTGTAT	TTTGAAGGTG	ATTTTACGGT
	1251	CTCGCCTGAA	AACAACGAAA	CGTGGCAAGG	CGCGGGCGTT	CATATCAGTG
	1301	AAGACAGTAC	CGTTACTTGG	AAAGTAAACG	GCGTGCCAAA	CGACCCGCTG
	1351	TCCAAAATCG	GCAAAGGCAC	GCTGCACGTT	CAAGCCAAAG	GGGAAAACCA
30	1401	AGGCTCGATC	AGCGTGGGCG	ACGGTACAGT	CATTTTGGAT	CAGCAGGCAG
	1451	ACGATAAAGG	CAAAAAACAA	GCCTTTAGTG	AAATCGGCTT	GNTCAGCGGC
	1501	AGGGGTACGG	TGCAACTGAA	TGCCGATAAT	CAGTTCAACC	CCGACAAACT
	1551	CTATTTGCGC	TTTCGCGGCG	GACGTTTGGA	TTTAAACGGG	CATTCCGCTT
	1601	CGTTCCACCG	TATTCAAAAT	ACCGATGAAG	GGGCGATGAT	TGNCNATCAT
35	1651	AATGCCACAA	CAACATCCAC	CGTTACCATT	ACAGGGAATG	AAAGTATTAC
	1701	ACAACCGAGT	GGTAAGAATA	TCAATAGACT	TAATTACAGC	AAAGAAATTG
	1751	CCTACACCGG	TTGGTTTGGC	GAGAAAGATA	CGACCAAAAC	GAACGGCGCG
	1801	CTCAACCTTG	TTTACCAGCC	CGCCGCAGAA	GACCCGACCC	NGCTGCTTTC
	1851	CGGCGGAACA	AATTTAAACG	GCAACATCAC	GCAAAACAAAC	GGCAAACTGT
40	1901	TTTTCAGCGG	CAGACCGACA	CCGCACGCCT	ACAATCATTT	AGGAAGCGGG
	1951	TGGTCAAAAA	TGGAAGGTAT	CCCACAAGGA	GAAATCGTGT	GGGACAACGA
	2001	CTGGATCNAC	CGCACGTTTA	AAGCGAAAAA	TTTCCATATT	CAGGGCGGGC
	2051	AGGCGGTGAT	TTCCCGCAAT	GTGCCAAAG	TGGAAGGCGA	TTGNCATTTG
	2101	AGCAATCAGC	CCCAAGCAGT	TTTTGGTGTC	GCACCGCATC	AAAGCCATAC
45	2151	AATCTGTACA	CGTTCGGACT	GGACNGGTCT	GACAAATTGT	GTCGAANAAA
	2201	NCATTACCGA	CGATAAAGTG	ATTGCTTCAT	TGACTAAGAC	NGACNTNAGC
	2251	GGCANTGTNA	GNCTNNCCNA	TNACGNTNNT	TNAAANCTCN	CNNGGCGNTGC
	2301	NNCACTNAAN	GGCAATCTTA	GTGCAATGG	CGATACACGT	TATACAGTCA
	2351	GCCACAACGC	CACCCAAAAC	GGCAACCTTA	GCCTCGTGGG	CAATGCCCAA
50	2401	GCAACATTTA	ATCAAGCCAC	ATTAAACGGC	AACNCATCGG	NTTCGGGCAA
	2451	TGCTTCATTT	AATCTAAGCA	ACAACGCCGC	ACAAAACGGC	AGTCTGACGC
	2501	TTTCCGACAA	CGCTAAGGCA	AACGTAAGCC	ATTCGCACT	CAACGGCAAT
	2551	GTCTCCCTAG	CCGATAAGGC	AGTATTCCAT	TTTGAAAACA	GCCGCTTTAC
	2601	CGGACAACTC	AGCGGCAGCA	AGGANACAGC	ATTACACTTA	AAAGACAGCG
55	2651	AATGGACGCT	GCCGTGAGGC	ACGGAATTAG	GCAATTTAAA	CCTTGACAAAC
	2701	GCCACCATTA	CACTCAATTC	CGCCTATCGC	CACGATGCTG	CAGGCGCGCA
	2751	AACCGGCAGN	GTGTGAGACA	CGCCGCGCGG	CCGTTGCGCG	CGTTCCCTAT
	2801	TATCCGTTAC	ACCGCCAAC	TCGGTAGAAT	CCCGTTTCAA	CACGCTGACG
	2851	GTAACCGGCA	AATTGAACNG	TCAAGGAACA	TTCCGCTTTA	TGTCGGAAC
60	2901	CTTCGGCTAC	CGAAGCGACA	AATTGAAGCT	GGCGGAAAGT	TCCGAAGGNA
	2951	CTTACACCTT	GGCGGTCAAC	AATACCGGCA	ACGAACCCGT	AAGCCTCGAT
	3001	CAATTGACGG	TAGTGGAAAG	GAAAGACAAC	AAACCGCTGT	CCGAAAACCT
	3051	TAATTTTACC	CTGCAAAACG	AACACGTCGA	TGCCGGCGCG	TGGCGTTACC
	3101	AACTCATCCG	CAAAGACGGC	GAGTTCCGCC	TGCATAATCC	GGTCAAGAA
65	3151	CAAGAGCTTT	CCGACAAACT	CGGCAAGGCA	GAAGCCAAAA	AACAGGCGGA
	3201	AAAAGACAAC	GCGCAAAAGC	TTGACGCGCT	GATTGCGGCC	GGGCGCGATG
	3251	CCGCCGAAAA	GACAGAAAGC	GTTGCCGAAC	CGGCCCGGCN	GGCAGGCGGG
	3301	GAAAAATGTC	GCATTATGCA	GGCGGAGGAA	GAGAAAAAAC	GGGTGACAGC
	3351	GGATAAAGAC	AGCGCNTTGG	CGAAACAGCG	CGAAGCGGAA	ACCCGGCCCG
	3401	NTACCACCGC	CTTCCCCCGC	GCCCGCNGCG	CCCGCCGGGA	TTTGCCGCAA
70	3451	CCGCAGCCCC	AACCGCAACC	TCAACCCCAA	CCGCAGCGCG	ACCTGATNAG
	3501	CCGTTATGCC	AATAGCGGTT	TGAGTGAATT	TTCCGCCACG	CTCAACAGCG
	3551	TTTTCGCCGT	ACAGGACGAA	TTGGACCGCG	TGTTTGCCGA	AGACCGCCCG

-365-

5  
 10  
 15

```

3601 AACGCNGTTT GGACAAGCNG CATCCGGNAC ACCAAACACT ACCGTTGCGA
3651 AGATTTCCGC GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC
3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTC GCACAACCGG
3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA
3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTCGAC ATCGGCATCA
3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TNTCAGACGG CATCGGAGGC
3901 AAAATCCGCC GCCGCGTGCT GCATTACGGC ATTACGGCAC GATACCGCGC
3951 CGGTTTCGGC GGATTTCGGCA TCGAACCGTA CATCGGCGCA ACGCGCTATT
4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC
4051 GGTCTTGCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTCATT
4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCTCTATA
4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG
4201 TCGCAGGATT TCGGCAAAAC CCGCAGTGCG GAATGGGGCG TAAACGCCGA
4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCCGCC AAAGNCCGC
4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 652>:

20  
 25  
 30  
 35  
 40  
 45

```

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYYRDAEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVS RNG
101 VAALVGDQYI VSAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPDN
151 SHPYNGDXHM PRLHKFVTD AEPVEMTSDMR GNTYSDKEY PERVRIGSGH
201 HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGF
251 MPIAGAAGDS GSPPMFYDKT NNKWLNLGVL QTGYPSYSGRE NGFQLIRKDW
301 FYDDIYRGDT HTVXFEP RSN GHFSFTSN NN GTGTVTETNE KVS NPKLKVQ
351 TVRLFDESLN ETDKEPVYAA GGVNQYRPR L NNGENLSFID YGNGKLILSN
401 NINQAGGLY FEGDFTVSPE NNETWQAGV HISEDSTVTW KVN GVANDRL
451 SKIKGTLHV QAKGENQGS I SVGDGTVILD QQADDKGKKQ AFSEIGLXSG
501 RGTVQLNADN QFNPKLYFG FRGGRDLN G HSLSFHRIQN TDEGAMIXXH
551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR
601 LNLVYQPAE DRTXLLSGGT NLNGNITQT N GKLFSSGRPT PHAYNHLGSG
651 WSKMEGIPQG EIVDNDWIX RTFKAENFHI QGGQAVISRN VAKVEGD XHL
701 SNHAQAVFGV APHQSH TICT RSDWTGLTNC VEXXITDDKV IASLT KTDXS
751 GXVXLXXXXX XXLXGXAXLX GNLSANGDTR YTVSHNATQN GNLSLVGNAQ
801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNGN
851 VSLADKAVFH FENSRTGQL SGSKXTALHL KDSEWTLPSG TELGNLNLDN
901 ATITLNSAYR HDAAGAQTX VSDTPRRRSR RSLLSVTPPT SVESR FNTLT
951 VNGKLNQGT FRFMSELFGY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD
1001 QLTVEVGKDN KPLSENLF T LQNEHVDAGA WRYQLIRKDG EFR LHPVKE
1051 QELSDKLGKA EAKKQAEKDN AQSLDALIAA GRDAEKTES VAEPARXAGG
1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TRPXTTAFPR ARXARRDL PQ
1151 PQPQPQPQPQ PQRDLSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR
1201 NAVWTSXIRX TKHYRSQDFR AYRQQTDLRQ IGMQKNLGSG RVGILFSHNR
1251 TENXFDDGIG NSARLAHGAV FGQYIGRFD IGISTGAGFS SGXLS DGI
1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQADY RYENVNIATP
1351 GLAFNRYRAG IKADYSFKPA QHXSTPYXS LSYTDAASGK VRTRVNTAVL
1401 AQDFGKTRSA EWGVNAEIKG FTLSXHAAA KGPQLEA QHS AGIKLGYRW*
  
```

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

50  
 55  
 60  
 65

```

              10      20      30      40      50      60
orfla.pep    MKTTDKRTTETHRKAPKTGRIREFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN
              |||
orfl-1       MKTTDKRTTETHRKAPKTGRIREFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN
              10      20      30      40      50      60

              70      80      90      100     110     120
orfla.pep    KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGDQYIVSAHNGGYN
              |||
orfl-1       KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGDQYIVSAHNGGYN
              70      80      90      100     110     120

              130     140     150     160     170     179
orfla.pep    NVDFGAEGXNPDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTD AEPVEMTSDM
              |||
orfl-1       NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTD AEPVEMTSYM
              130     140     150     160     170     180
  
```

-366-

		180	190	200	210	220	230
	orf1a.pep	RGNTYSDKEKYPERVIRIGSGHHYRYDDDKHGDL--SYSGA---WLIGGNTHMQGWGNN					
	orf1-1	DGRKYIDQNNYPDRVRIGAGRQYWRSEDEFPNNRESSYHIASAYSWLVGNGTFAQNGSGG					
5		190	200	210	220	230	240
	orf1a.pep	240	250	260	270	280	290
	orf1-1	GVXSLSGD-VRHANDYGPMPFIAGAAGDSGSPMFIYDKTNKWLNGVLQTGYPYSGRENG					
10		250	260	270	280	290	
	orf1a.pep	GTVNLGSEKIKHS-PYGFLPTGGSGFGDSGSPMFIYDAQKQKWLINGVLQTGNPFYIGKSNG					
	orf1-1	GTVNLGSEKIKHS-PYGFLPTGGSGFGDSGSPMFIYDAQKQKWLINGVLQTGNPFYIGKSNG					
		300	310	320	330	340	350
	orf1a.pep	FQLIRKDWFYDDIYRGDTHTVXFEPERSNGHFSFTSNNGTGTVTETNEKVSNP-KLKVQT					
15		300	310	320	330	340	350
	orf1a.pep	FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT					
	orf1-1	FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT					
		360	370	380	390	400	410
	orf1a.pep	VRLFDESINETDKEPVY-AAGGVNQYRPRLNNGENLSFIDYGNGLILSNINQAGAGGLY					
20		360	370	380	390	400	410
	orf1a.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLY					
	orf1-1	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLY					
		420	430	440	450	460	470
	orf1a.pep	FEGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGSI					
25		420	430	440	450	460	470
	orf1a.pep	FQGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGSI					
	orf1-1	FQGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGSI					
		480	490	500	510	520	530
	orf1a.pep	SVGDGTVILDQQADDKGKKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG					
30		480	490	500	510	520	530
	orf1a.pep	SVGDGTVILDQQADDKGKKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG					
	orf1-1	SVGDGTVILDQQADDKGKKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG					
		540	550	560	570	580	590
	orf1a.pep	HSLSFHRIQNTDEGAMIXXHNATTTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFG					
35		540	550	560	570	580	590
	orf1a.pep	HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDIAT-TGNN-NSLDSKKEIAYNGWFG					
	orf1-1	HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDIAT-TGNN-NSLDSKKEIAYNGWFG					
		600	610	620	630	640	650
	orf1a.pep	EKDTTKTNGRLNLVYQPAEDRTXLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSG					
40		600	610	620	630	640	650
	orf1a.pep	EKDTTKTNGRLNLVYQPAEDRTLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDH					
	orf1-1	EKDTTKTNGRLNLVYQPAEDRTLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDH					
		660	670	680	690	700	710
	orf1a.pep	WSKMEGIPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKVEGDHXLNHAQAVFGV					
45		660	670	680	690	700	710
	orf1a.pep	WSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHXLNHAQAVFGV					
	orf1-1	WSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHXLNHAQAVFGV					
		720	730	740	750	760	770
	orf1a.pep	APHQSHTICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVXLXXXXXXLXGXAXLX					
50		720	730	740	750	760	770
	orf1a.pep	APHQSHTICTRSDWTGLTNCVEKTTITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLN					
	orf1-1	APHQSHTICTRSDWTGLTNCVEKTTITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLN					
		780	790	800	810	820	830
	orf1a.pep	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSNGNASFNLSNNAQNG					
55		780	790	800	810	820	830
	orf1a.pep	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNG					
	orf1-1	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNG					
		840	850	860	870	880	890
	orf1a.pep	SLTSLDNAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSG					
60		840	850	860	870	880	890
	orf1a.pep	SLTSLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					
	orf1-1	SLTSLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					
		840	850	860	870	880	890
	orf1a.pep	SLTSLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					
65		840	850	860	870	880	890
	orf1a.pep	SLTSLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					
	orf1-1	SLTSLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					
		840	850	860	870	880	890
	orf1a.pep	SLTSLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					
70		840	850	860	870	880	890
	orf1a.pep	SLTSLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					
	orf1-1	SLTSLSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					

-367-

		900	910	920	930	940
	orf1a.pep	TELGNLNLNDNATITLNSAYRHDAAQAQTGXVSDTPRRRSRRS---	LLSVTPPTS	SVESRFN		
	orf1-1	TELGNLNLNDNATITLNSAYRHDAAQAQTGSATDAPRRRSRRSRRSLLSVTPPTS	SVESRFN			
5		900	910	920	930	940
	orf1a.pep	950	960	970	980	990
	orf1-1	950	960	970	980	990
10		960	970	980	990	1000
	orf1a.pep	1010	1020	1030	1040	1050
	orf1-1	1010	1020	1030	1040	1050
15		1020	1030	1040	1050	1060
	orf1a.pep	1070	1080	1090	1100	1110
	orf1-1	1070	1080	1090	1100	1110
20		1080	1090	1100	1110	1120
	orf1a.pep	1130	1140	1150	1160	1170
	orf1-1	1130	1140	1150	1160	1170
25		1140	1150	1160	1170	1180
	orf1a.pep	1190	1200	1210	1220	1230
	orf1-1	1190	1200	1210	1220	1230
30		1200	1210	1220	1230	1240
	orf1a.pep	1250	1260	1270	1280	1290
	orf1-1	1250	1260	1270	1280	1290
35		1260	1270	1280	1290	1300
	orf1a.pep	1310	1320	1330	1340	1350
	orf1-1	1310	1320	1330	1340	1350
40		1320	1330	1340	1350	1360
	orf1a.pep	1370	1380	1390	1400	1410
	orf1-1	1370	1380	1390	1400	1410
45		1380	1390	1400	1410	1420
	orf1a.pep	1430	1440	1450		
	orf1-1	1430	1440	1450		
50		1440	1450			
55						
60						

Homology with adhesion and penetration protein hap precursor of *H. influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

	orf1	23	FXAAYLAICLSFGILPQAWAGHTYFGINYYQYRDAENKKGKFAVGAKDIEVYNKKGELVG	82
	hap	6	FRLNFLTACVSLGIASQAWAGHTYFGIDYQYRDAENKKGKFTVGAKNIEVYNKEGQLVG	65
	orf1	83	KSMTKAPMIDFSVVSRRNGVAALVGVQYIVSVAHNGGYNVDFGAEGXNIXDQXRXTYKIV	142
	hap	66	TSMTKAPMIDFSVVSRRNGVAALVGDQYIVSVAHNGGYNVDFGAEGRN-PDQHRFTYQIV	124

-368-

5  
10  
15  
20

```

orfl 143 KRNNYKAGTKGHPYGGDYHMPRLHKXVTD AEPVEMTSYMDGRKYIDQNNYPDRVRIGAGR 202
      KRNNY+A + HPY GDYHMPRLHK VT+AEPV MT+ MDG+ Y D+ NYP+RVRIG+GR
hap 125 KRNNYQAWERKHPYDGDYHMPRLHKFVTEAEPVGMTTMDGKVYADRENYPERVRIGSSGR 184

orfl 203 QYWRSEDEPNNRESSYHIA----- 222
      QYWR+D+DE N SSY+++
hap 185 QYWRDDEETNVHSSYYVSGAYRYLTAGNTHQTSGNGNGTVNLSGNNVSPNHYGPLPTG 244

orfl 223 -----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVF 277
      SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF
hap 245 GSKGDSGSPMFIYDAKKKQWLINAVLQTGHPPFGRNGNGFQLIREEFYNEVLAVDTPSVF 304

orfl 278 --YEPRQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA 334
      Y P NG YSF +N+GTGK+ + + + TV+LFN SL++TA+E V A
hap 305 QRYIPPINGHYSFVSNNDGTGKLTLTRPSKDGSKAKSEVGTVKLFNPSLNQTAKEHV-KA 363

orfl 335 AGGVNSYRPRLNNGENISFIDEKGELILTSNINQGAGGLYFQGDFTV-SPENNETWQGA 393
      A G N Y+PR+ G+NI D+GKG L + +NINQGAGGLYF+G+F V +NN TWQGA
hap 364 AAGYNIYQPRMEYGKNIYLGDDQGGKTLTIENNINQGAGGLYFEGNFVVKGKQNNITWQGA 423

orfl 394 GVHISEDSTVTWKVNGVANDRLSKIGKGT 423
      GV I +D+TV WKV+ NDRLSKIG GTL
hap 424 GVSIGQDATVEWKVHNPENDRLSKIGIGT 453

```

25 Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:

30  
35  
40

```

Orfl 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSDHAVQNGSLTLS 98
      DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS
hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTLDHDSQFTLSNNATQTGNIKLS 792

orfl 99 GNAKANVSHSALNGNVSLADKAVHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN 158
      +A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N
hap 793 NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDTVTLENATWTMPSDTTLQN 852

orfl 159 LNLDNATITLNSAYRHDAAQAQTGSATDAPXXXXXXXXXXLLXVTPPTSVESRENTLTVN 218
      L L+N+T+TLNSAY + S+ +AP L T PTS E RENTLTVN
hap 853 LTLNNSTVTLNSAY-----SASSNNAPRHRS-----LETETTP TSAEHRNTLTVN 899

orfl 219 GKLNGQGTFRFMSELFGYRSDKLKLAESSEGTYT LAVNNTGNEPASLEQLTVVEGKDNKP 278
      GKL+GQGTFF S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP
hap 900 GKLSGQGTFFQFTSSLFYKSDKLKLSNDAEGDYTL SVRNTGKEPVTLEQLTLIESLDNKP 959

orfl 279 LSENLFNLQNEHVDAGA 296
      LS+ L FTL+N+HVDAGA
hap 960 LSDKLKFTLENDHVDAGA 977

```

45 Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:

50  
55  
60  
65

```

Orfl 1 LDRVFAEDRRNAVWTS GIRD TKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRV GILFSHNR 60
      LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R
hap 1135 LDRLEVDQAQSAVWTNIAQDKRRYDSDAFRAYQQKTNLRLQIGVQKALANGRIGAVFSHSR 1194

orfl 61 TENTFDDGIGNSARLAHGA VFGQY GIDRFYXXXXXXXXXXXXXXXXXIGXKXRRRVLHYG 120
      ++NTFD+ + N A L + F QY K R+ ++YG
hap 1195 SDNTFDEQVKNHATLTMMSGFAQYQWGD LQFGVNVGTGISASKMAEEQSRKIHRKAINYG 1254

orfl 121 IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA 180
      + A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P
hap 1255 VNASYQFRLGQLGIQPYFGVNRYPIERENYQSEEV RVKTPSLAFNRYNAGIRVDYTFPT 1314

orfl 181 QHISITPYLSLSYTD AASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA 240
      +IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +
hap 1315 DNISVKPYFFVNYVDVSNANVQT VNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKS 1374

orfl 241 KGPQLEA QHSAGIKLGYRW 259
      +G QL Q + G+KLGYRW
hap 1375 QGSQLGKQQNVGVKLG YRW 1393

```

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orf1.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN	60
	orf1ng	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN	60
10	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVG VQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALAGDQYIVSVAHNGGYN	120
15	orf1.pep	NVDFGAEGXNIXDQXRXYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSY	180
	orf1ng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSY	179
20	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGGNTFAQNGSG	239
25	orf1.pep	-----GSPMFIYDAQKQKWLINGVLQTGNPYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGF LPTGGSGFGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNG	289
30	orf1.pep	<u>FQLVRKDWFYDEIFAGDTHSVFYEP</u> RQNGKYSFNDNNGTGKINAKHEHNSLPNRLKTRT	315
	orf1ng	<u>FQLVRKDWFYDEIFAGDTHSVFYEP</u> HQNGKYFFNDNNAGAKIDAKHKHYSLPYRLKTRT	359
35	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDKGGELILTSNINQAGAGGLY	
40	orf1.pep	FQGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNNETWQGAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
45	orf1.pep	// DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
50	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLS DHA	803
	orf1ng	TFNGNL-VQAETRTIRLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
55	orf1.pep	VQNGSLTSLGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWT	863
	orf1ng	VQNGSLTSLDNAKANVSHSALNGNVSLADKAVFHFENS RFTGKISGGKDTALHLKDSEWT	893
60	orf1.pep	LPSGXLGNLNDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRLXVTPPTSVE	923
	orf1ng	LPSGTELGNLNDNATITLNSAYRHDAAGAQTGSAADAPRRRSRRS---LLSVTPPTS AE	950
65	orf1.pep	SRFNTLTVNGKLNQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLT	983
	orf1ng	SRFNTLTVNGKLNQGTFRFMSELFGYRSGKLKLAESSEGTYTLAVNNTGNEPVSLEQLT	1010
70	orf1.pep	VVEGKDNKPLSENLFNTLQNEHVDAGAW	1011
	orf1ng	VVEGKDNTPLSENLFNTLQNEHVDAGAWRYQLIRKDG EFR LHPVKEQELSDKLGKAGET	1070
75	orf1.pep	// LDRVFAEDRRNAVWTS GIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTS GIRDTKHYRSQDFR	1239
80	orf1.pep	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGA VFGQYGIDRFY	1271
	orf1ng	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTGNTFDDGIGNSARLAHGA VFGQYGIGRFD	1299

```

orfl.pep      IGISAGAGFSSGSLSDGIGXKXRRRLVHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1331
               |||||
orfl.ng       IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1359

5  orfl.pep      RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVL 1391
               |||||
orfl.ng       RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVL 1419

10 orfl.pep      AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGPQLEAQSAGIKLGYRW 1440
               |||||
orfl.ng       AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGPQLEAQSAGIKLGYRW 1468

```

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

```

1  ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCTAA
15 51 AACCGGCCGC ATCCGCTTCT CGCCCGCTTA CTTAGCCATA TGCCTGTCTG
   101 TCGGCATTCT GCCCCAAGCC CGGGCGGGAC ACACTTATTT CGGCATCAAC
   151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
   201 GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAAT
   251 CGATGACGAA AGCCCCGATG ATTGATTTTT CTGTGGTATC GCGTAACGGC
   301 GTGGCGGCAT TGGCGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
20 351 CGGCTATAAC AATGTTGATT TTGGTGC GGAAGCAAT CCCGATCAGC
   401 ACCGCTTTTC TTACCAAAAT GTGAAAAGAA ATAATATAAA AGCAGGGACT
   451 AACGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCTT TGCACAAATT
   501 TGTCACAGAT GCAGAACCTG TTGAGATGAC CAGTTATATG GATGGGTGGA
   551 AATACGCTGA TTAAATAAAA TACCCTGATC GTGTTGCAAT CGGAGCAGGC
25 601 AGACAATATT GCGGCTCTGA TGAAGACGAA CCCAATAACC GCGAAAGTTC
   651 ATATCATATT GCAAGCGCAT ATTCTTGGCT CGTCGGTGGC AATACCTTTG
   701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG CGAAAAAATT
   751 AAACATAGCC CATATGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
   801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
30 851 ATGGGGTATT GCAAAACAGG AACCCTATA TAGGAAAAAG CAATGGCTTC
   901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
   951 CCATTAGTA TTCTACGAAC CACATCAAAA TGGGAAATAC TTTTTTAACG
1001 ACAATAATAA TGGCGCAGGA AAAATCGATG CCAACATAAA ACACATTATCT
1051 CTACCTTATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
35 1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCCTT TATTGACAAA
1201 GGAAAAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAACAGC
40 1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGATG GCAGTACCGT TACTTGAAAA
1351 GTAAACGGCG TGGCAAAACG CCGCCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCTGGTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGAGC
1451 GATAAGTCAT CTTAGATCAG CAGGCGGACG ATCAAGGCAA AAAACAGGCC
1501 TTTAGTGAAG TCGGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC
1551 CGATAATCAG TTCAACCCCG ACAAATCTTA TTTCCGCTTT CGCGGCGGAC
45 1601 GTTTGGATTT GAACGGGCAT TCGCTTTCGT TCCACCGCAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTACTAC AACCAGCAAT AACACAACCT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
50 1801 GCAACCAAAA CGAACGGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC
1851 GGATCGCACT TTAATGCTTT CCGGCGGAAC AAATTTAAAC GGCAATATCA
1901 CGCAACAAAA CGGCAAACTG TTTTTCAGCG GCAGACCGAC ACCGCACGCC
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCACAAGG
2001 AGAAATCGTG TGGGACAACG ATTGGATCGA CCGCACATTT AAAGCGGAAA
55 2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCCGCAA TGTTGCCAAA
2101 GTGGAAGGCG ATTGGCATT T AAGCAATCAC GCCAAGCAG TTTTCGGTGT
2151 CGCACC GCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAGTTG TACCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCCG ATCAGCTCA
60 2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCTT AGTGCAGGCG
2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCAAAAA CGGCAACCTC
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACACGCGG
2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AAACGTAAGC
2551 CATTCCGCAC TCAACGGCAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
65 2601 TTTTGAAAAC AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACCG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCGGG CACGGAATTA
2701 GGCAATTAA ACCTTGACAA CGCCACCATT ACACTCAATT CCGCCTATCG
2751 ACACGATGCG GCAGGCGCGC AAACCGGCAG TGCCGCGAGT GCGCCGCGC
2801 GCCGTTTCGC CCGTTCCCTA TTATCCGTTA CGCCGCCAAC TTCGGCAGAA
70 2851 TCCCGTTTCA ACACGCTGAC GGTAACGGGC AAATTGAACG GTCAGGGAAC

```



-371-

2901 ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC  
 2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCCGC  
 3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGGAG GAAAAGACAA  
 3051 CACACCGCTG TCCGAAAATC TTAATTTTAC CTGCaAAc gaacacgtcg  
 5 3101 atgccggcgc atggCGTTAT CAGCTTATCC gcaaagacgG CGAGTTCCgc  
 3151 CTGCATAATC CGGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGGc  
 3201 gggagaaACA GAggccgccT TGACGGCAAA ACAGGCacaA CTTGCCGCCA  
 3251 AACaacaggc ggaaaAAGAC AACgcgcaaa gccttgAcgc gctgattgcg  
 10 3301 gCggggcgca atgccaccga AAAGGCagaa agtgttgccg aaccgGCCCCG  
 3351 GCAGGCAGGC GGGGAAAAtg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA  
 3401 AACGGGTGCA GCGGATATAA GACACCGCCT TGGCGAAACA GCGCGAAGCG  
 3451 GAAACCCGGC CGGCTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCCG  
 3501 GGATTTGCCG CAACCGCAGC CCAACCGCA ACCCAACCG CAGCGCGACC  
 3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC  
 15 3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA  
 3651 CCGCCGCAAC GCGTTTGGGA CAAGCGGCAT CCGGGACACC AAACACTACC  
 3701 GTTCGCAAGA TTTCCGCGCC TACCGCCAAC AAACCGACCT GCGCCAAATC  
 3751 GGTATGCAGA AAAACCTCGG CAGCGGGCGC GTCGGCATCC GTTTTTCCGA  
 20 3801 CAACCGGACC GGAAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC  
 3851 TTGCCCACGG TGCCGTTTTC GGGCAATACG GCATCGGCAG GTTCGACATC  
 3901 GGCATCAGCG CGGGCGCGGG TTTTAGTAGC GGCAGCCTTT CAGACGGCAT  
 3951 CAGAGGCAAA ATCCGCCGCC GCGTGCTGCA TTACGGCATT CAGGCAAGAT  
 4001 ACCGCGCAGG TTTGCGCGGA TTCGGCATCG AACCGCATAT CGGCGCAACG  
 4051 CGCTATTTTCG TCCAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC  
 25 4101 CACCCCGGGC CTTCGCATCA ACCGCTACCG CGCGGGCATT AAGGCAGATT  
 4151 ATTCATTCAA ACCGCGCGCA CACATTTCCA TCACGCCTTA TTTGAGCCTG  
 4201 TCCTATACCG ATGCCGCTTC CGGCAAGTTC CGAACCGCGC TCAATACGCG  
 4251 CGTATTGGCG CAGGATTTTCG GCAAAACCCG CAGTGCGGAA TGGGGCGTAA  
 4301 ACGCCGAAAT CAAAGGTTTC ACGCTGTCCC TCCACGCTGC CGCCGCCAAG  
 30 4351 GGGCCGCAAT TGAAGCGCA GCACAGCGCG GGCATCAAA TAGGCTACCG  
 4401 CTGGTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA RAGHTYFGIN  
 35 51 YQYYRDFAEK KGFFAVGAKD IEVYNKRGEL VGKSMTKAPM IDFSVSVSRNG  
 101 VAALAGDQYI VSAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNNYKAGT  
 151 NGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG  
 201 RQYWRSEDE PNNRESSYHI ASAYSWLVGG NTFQNGSGG GTVNLGSEKI  
 251 KHSPYGFLLPT GGSFGDSGSP MFIYDAQKQK WLINGVLOTG NPYIGKSN  
 301 QLVRKDWFYD EIFAGDTHSV FYEPHONGKY FFNDNNNGAG KIDAKHKHYS  
 40 351 LPYRLKTRTV QLFNVLSLET AREPVYHAAG GVNSYRPRLN NGENISFIDK  
 401 KGELILTSN INQAGGLYF EGNFTVSPKN NETWQAGVH ISDGSTVTWK  
 451 VNGVANDRLS KIGKGTLLVQ AKGENQGSVS VGDGKVILDQ QADDQGGKQA  
 501 FSEIGLVSGR GTVOLNADNQ FNPDKLYFGF RGRRLDLNGH SLSFHRINT  
 551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNGWFGEKD  
 45 601 ATKTNGLNL NYPPEADRT LLSGGTNLN GNITQTNGL FFSGRPTPHA  
 651 YNHLGSGWSK MEGIPQGEIV WDNDWIDRTF KAENFHIQGG QAVVSRNVAK  
 701 VEGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS  
 751 LSKTDVRGNV SLADHAHLNL TGLATFNGNL VQAEITRTRL RANATQNGNL  
 801 SLVGNAQATF NOATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS  
 50 851 HSAALNGNVS ADKAVFHFEN SRFTGKISGG KDTALHLKDS EWTLPSTGTEL  
 901 GNLNLDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRSL LSVTPPTSAE  
 951 SRFNTLTVNG KLNQGGTFRF MSELFYGRSG KLKLAESSEG TYTLAVNNTG  
 1001 NEPVSLEQLT VVEGKDNTPL SENLNFTLQN EHVDAWAWRY QLIRKDGFR  
 1051 LHPVKEQEL SDKLGKAGET EAALTAKQAQ LAAKQQAQEKD NAQSLDALIA  
 55 1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKRVQADK DTALAKQREA  
 1151 ETRPATTAFF RARRARRDLF QPQPQPQPQ QRDILISRYN SGLSEFSATL  
 1201 NSVFAVQDEL DRVFAEDRRN AVWTSIGRDT KHYRSQDFRA YRQQTDLRQI  
 1251 GMQKNLGSGR VGILFSHNRT GNTFDDGIGN SARLAHGAVF GQYIGIRFDI  
 1301 GISAGAGFSS GSLSDGIRGK IRRRLHYGI QARYRAGFGG FGIEPHIGAT  
 60 1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSFKPAQ HISITPYLSL  
 1401 SYTDAASGVK RTRVNTAVIA QDFGKTRSAE WGVNAEIKGF TSLHAAAK  
 1451 GPQLEAQHSA GIKLGYRW\*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

-372-

		10	20	30	40	50	60
	orf1-1.pep	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN					
5	orf1ng-1	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN					
		10	20	30	40	50	60
	orf1-1.pep	70	80	90	100	110	120
10	orf1ng-1	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGDQYIVSVAHNNGGYN					
		70	80	90	100	110	120
	orf1-1.pep	130	140	150	160	170	180
15	orf1ng-1	NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAEPVEMTSYM					
		130	140	150	160	170	180
	orf1-1.pep	190	200	210	220	230	240
20	orf1ng-1	DGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGNTFAQNGSGG					
		190	200	210	220	230	240
25	orf1-1.pep	250	260	270	280	290	300
	orf1ng-1	GTVNLGSEKIKHSPYGF LPTGGSGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF					
30		250	260	270	280	290	300
	orf1-1.pep	310	320	330	340	350	360
	orf1ng-1	QLVRKDWFYDEIFAGDTHSVFYEPHQNGKYSFNDNNGTGKINAKHEHNSLPNRLKTRTV					
35		310	320	330	340	350	360
	orf1-1.pep	370	380	390	400	410	420
40	orf1ng-1	QLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEKGKELILTSNINQAGAGGLYF					
		370	380	390	400	410	420
	orf1-1.pep	430	440	450	460	470	480
45	orf1ng-1	QGDFTVSPENNETWQAGVHISEDSTVTWKVNGVANDRLSKIGKGTILHVQAKGENQGSIS					
		430	440	450	460	470	480
	orf1-1.pep	490	500	510	520	530	540
50	orf1ng-1	VGDGTVILDQQADDQKKQAFSEIGLVSGRGT VQLNADNQFNPDKLYFGRGRLDLNGH					
		490	500	510	520	530	540
55	orf1-1.pep	550	560	570	580	590	600
	orf1ng-1	SLSFHRIQNTDEGAMIVNHNQDKESTVTTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKD					
60		550	560	570	580	590	600
	orf1-1.pep	610	620	630	640	650	660
	orf1ng-1	TTKTNGRLNLVYQPAEDRTL LLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQ					
65		610	620	630	640	650	660
	orf1-1.pep	670	680	690	700	710	720
70	orf1ng-1	KEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLNSHAQAVFGVAPH					
		670	680	690	700	710	720

10

15

20

25

30

35

40

45

50

55

65

```

1370      1380      1390      1400      1410      1420
orfl-1.pep  AGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI
              |||||
orflng-1    AGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI
1380      1390      1400      1410      1420      1430

```

-374-

orfl-1.pep 1430 1440 1450  
 KGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX  
 |||||  
 orflng-1 1440 1450 1460  
 KGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX

In addition, ORF1ng shows 55.7% identity with hap protein (P45387) over a 1455aa overlap:

SCORES Init1: 1104 Initn: 4632 Opt: 2680  
 Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

10	orflng-1.pep	10 20 30 40 50 60	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDFAE
15	p45387	10 20 30 40	MKKTVFRNLNLTACISLGIVSQAWAGHTYFGIDYQYYRDFAE
20	orflng-1.pep	70 80 90 100 110 120	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALAGDQYIVSVAHNGGYN
	p45387	50 60 70 80 90 100	KGKFTVGAQNIKVVYNKQGLVGTSMTKAPMIDFSVVS RNGVAALVENQYIVSVAHNVGYT
25	orflng-1.pep	130 140 150 160 170 180	NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSYM
	p45387	110 120 130 140 150 160	DVDFGAEGNNPDQHRFTYKIVKRNNYKGD-NLHPYEDDYHNPRLHKFVTEAAPIDMTSNM
30	orflng-1.pep	190 200 210 220 230 240	DGWKYADLNKYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGNGTFAQNGSGG
	p45387	170 180 190 200 210	NGSTYSDRTKYPERVRIGSGRQFWRNDQDKG-----QVAGAYHYLTAGTNHNRQAGAN
35	orflng-1.pep	250 260 270 280 290 300	GTVNLGSEKIKHSPYGFLLPTGGSGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF
40	p45387	220 230 240 250 260 270	GYSYLGSDVRKAGEYGPLPIAGSKGDSGSPMFIYDAEKQKWLINGILREGNPFEGKENG
45	orflng-1.pep	310 320 330 340 350 360	QLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHSLPYRLKTRTV
	p45387	280 290 300 310 320	QLVRKSYF-DEIFERDLHTSLYTRAGNVYITISGNDNGQGSITQKS---GIPSEIK---I
50	orflng-1.pep	370 380 390 400 410 419	QLFNVSLSETAREPVYHAA-GGVNSYRPRLNNGENISFIDKKGKELILTSNINQAGAGLY
	p45387	330 340 350 360 370 380	TLANMSLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDQKQGSILFASDINQAGAGLY
55	orflng-1.pep	420 430 440 450 460 470 479	FEGNFTVSPKNNETWQAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV
	p45387	390 400 410 420 430 440	FEGNFTVSPNSNQTWQAGAGIHVSENSTVTWKVNGVEHDLRLSKIGKGTLLHVQAKGENKGI
60	orflng-1.pep	480 490 500 510 520 530 539	SVGDGKIVLDQADDQKKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG
	p45387	450 460 470 480 490 500	SVGDGKIVILEQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKFYFGFRGGRDLNG
65	orflng-1.pep	540 550 560 570 580 590	HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITT-TGNN-NNLDSKKEIAYNGWFG
70	p45387	510 520 530 540 550 560	HSLTFKRIQNTDEGAMIVNHNNTQAANVTITGNESIVLPNGNNINKLDYRKEIAYNGWFG

	orf1ng-1.pep	600	610	620	630	640	650
		EEKDATKTN	GRNLNLNQ	PEEADRTLL	LSGGTNLNG	NIQTNGKLF	FFSGRPTPHAYNHLGSG
		:		:		:	:
5	p45387	ETDKNKHNG	RNLNLIYK	PTTEDRTLL	SSGGTNLKG	DDITQTKGK	LFSSGRPTPHAYNHLNKR
		570	580	590	600	610	620
	orf1ng-1.pep	660	670	680	690	700	710
		WSKMEGIPQ	GEIVWDND	WIDRTFKA	ENFHIQGG	QAVVSRNV	AKVEGDWHLSNHAQAVFGV
		:		:	:	:	:
10	p45387	WSEMEGIPQ	GEIVWDHD	WINRTFKA	ENFQIKGS	SAVVSRNV	SSIEGNWTVSNNANATFGV
		630	640	650	660	670	680
	orf1ng-1.pep	720	730	740	750	760	770
		APHQSHTIC	TRSDWTGL	TSCTEKTI	TDDKVIAS	LSKTDIRGN	VSADHAHLNLTGLATIN
		:		:	:	:	:
15	p45387	VPNQQTIC	TRSDWTGL	TTCQKVDL	TDTKVINI	SPKTQING	SINLTNDATANVKGLAKLN
		690	700	710	720	730	740
	orf1ng-1.pep	780	790	800	810	820	830
		GNLSAGGD	THYTVTRN	ATQNGNLS	VLGVNAQA	TFNQATLNG	NTSASDNASFNLSSNAVQNG
		:	:	:	:	:	:
20	p45387	GNVTL-----	-----	-----	-----	-----	-----TNHSQFTLSNNATQIG
		750					760 770
	orf1ng-1.pep	840	850	860	870	880	890
		SLTLSDNA	KANVSHS	ALNGNVSL	ADKAVFHF	ENSFRFTG	KISGGKDTALHLKDSEWTLPSG
			:	:	:     :	:     :	:     :
30	p45387	NIRLSDN	STATVDN	ANLNGNV	HLTDSAQF	SLKNSHF	SHQIQDGKGTTVTLLENATWTMPSD
			780	790	800	810	820 830
	orf1ng-1.pep	900	910	920	930	940	950
		TELGNLNL	DNATITL	NSAYRHD	AAGAQTGS	AADAPRR	RRSRLSSVTPTSAESRFNTLT
			:	:	:	:	
35	p45387	TTLQNLT	LNNSTIT	NSAY-----	SASSNNT	PRRRS---	LETETTPTSAEHRFNTLT
			840		850	860	870
	orf1ng-1.pep	960	970	980	990	1000	1010
		VNGKLNQ	QGTFRF	MSELFGY	RSGLKLA	ESSEGT	TYTLAVNNTGNEPVSLEQLTVVEGKDN
40	p45387	VNGKLSQ	QGTFFQ	TSSFLGY	KSDKLK	LSNDAEG	GDYILSVRNTGKEPETLEQLTLVESKDN
		880	890	900	910	920	930
	orf1ng-1.pep	1020	1030	1040	1050	1060	1070
		TPLSENLN	FTLQNE	HVDAGAW	RYQLIRK	DGEFRL	HPVKEQELSDKLGKAGETEAAALTAK
45	p45387	QPLSDKL	KFTLE	NDHVDAG	ALRYKL	VKNDDGE	FRLHNPKEQELHNDLVRAEQAERTLEAK
		940	950	960	970	980	990
	orf1ng-1.pep	1080	1090	1100	1110	1120	1130
		QAQLAAK	QQAQEK	DNQSLD	ALIAAG	RNAT-EKA	ESVAEPARQAGGENAGIMQAEKKRV
		:	:	:	:	:	:
50	p45387	QVEPTAK	TQTGE	PKVRSR	RAARAF	PDTLPDQ	SLLNALEAKQAE-LTAETQKS
		1000	1010	1020	1030	1040	1050
	orf1ng-1.pep	1140	1150	1160	1170	1180	1190
		QADK---	DTALAK	QREAE	TRPATT	AFPRARR	ARRD-LPQPQPQPQPQORDLISRYANS
		:   :	:   :	:   :	:   :	:   :	:   :
60	p45387	RSKRAVF	SDPLLD	QSLFALE	AALEVI	DAPQQSE	KDRLAQEEAEKQ-RKQKDLSISRYNSA
		1060	1070	1080	1090	1100	1110
	orf1ng-1.pep	1200	1210	1220	1230	1240	1250
		LSEFSAT	LSNVFA	VQDELDR	VFAEDRR	NAVWTS	GIRDTKHYRSQDFRAYRQQ-TDLRQIG
65	p45387	LSELSAT	VNSML	SVQDEL	DLRFVD	QAQSAV	WTNIAQDKRRYSDAFRAYQQQKTNLRQIG
		1120	1130	1140	1150	1160	1170
	orf1ng-1.pep	1260	1270	1280	1290	1300	1310
		MQKNLGS	GRVGIL	FSHNRT	GTNTFDD	GIGNSAR	LAHGAVFGQYQIGRFDIGISAGAGFSSG
		:	:	:	:		

-376-

		1320	1330	1340	1350	1360	1370
	orf1ng-1.pep	SLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGL					
	p45387	:     :    :   :   :     :  :  :  :					
5		1240	1250	1260	1270	1280	1290
	orf1ng-1.pep	AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEW					
	p45387	:      :  :  :   :     :     :         :					
10		1300	1310	1320	1330	1340	1350
	orf1ng-1.pep	GVNAEIKGFTLSLHAAAAGPQLEAQHSAGIKLGYRWX					
	p45387	:       :  :  :       :  :					
15		1360	1370	1380	1390		

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 20 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

	1	..AAGGTGTGGC	AATTTGTGCA	AGA.CCGCTG	CGTGCCGTCG	TGCCTGCCGA
	51	CAGTTTGTGA	CCGACCGCGC	AAAAATTGAA	CCTGTTTAAG	GCGGGTGGCG
	101	CAACCATTTT	GTTTTATGAA	GATCAAAATG	TCGTCAAAGG	TTTGCAAGGAG
25	151	CAGTTCCTTG	CTTATGCCCG	TAAC'TCCCC	GTTTGGGCGG	ATCAGGCAAA
	201	CGCGATGGTG	CAGTATGCCG	TTTGGACGAC	ACTTGCCGCG	GTCGGCGTAG
	251	GTGCAACCT	GCAACATTAC	AATCCCTTGC	CCGATGCGGC	GATTGCCAAA
	301	GCGTGGGAATA	TCCCCGAAAA	CTGTTGTGTTG	CGCGCACAAA	TGGTTATCGG
	351	CGGTATTGAA	GGGGCGGCAG	GTGAAAAGAC	CTTTGAACCC	GTTGCAACAAC
30	401	GTTTGAAGT	GTTCCGCGCA	TAA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

	1	..KVVQFVEXPL	RAVVPADSFE	PTAQKLNLFK	AGAATILFYE	DQNVVKGLQE
	51	QFPAYAAANF	VWADQANAMV	QYAVWTTLAA	VGVGANLQHY	NPLPDAAIAK
	101	AWNIPENWLL	RAQMVIGGIE	GAAGEKTFEP	VAERLKVFGA	*

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

	1	..CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAAAAT
	51	GAACCTGTTT	AAGGCGGGTG	CGGCAACCAT	TTTGTTTTAT	GAAGATCAAA
	101	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATGC	CGCTAACTTC
	151	CCCGTTTGGG	CGGATCAGGC	AAACGCGATG	GTGCAGTATG	CCGTTTGGAC
40	201	GACACTTGCC	GCGGTCGGCG	TAGGTGCAAA	CCTGCAACAT	TACAATCCCT
	251	TGCCCCGATG	GGCGATTGCC	AAAGCGTGGA	ATATCCCCGA	AAACTGGTTG
	301	TTGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGCGCG	CAGGTAAAAA
	351	GACCTTTGAA	CCCGTTGCAG	AACGTTTGAA	AGTGTTCGGC	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

45	1	..LRVAVPADSF	EPTAQKLNLF	KAGAATILFY	EDQNVVKGLQ	EQFPAYAAANF
	51	PVWADQANAM	VQYAVWTTLA	AVGVGANLQH	YNPLPDAAIA	KAWNIPENWL
	101	LRAQMVIGGI	EGAAGEKTFE	PVAERLKVFG	A*	

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

```

5      orf6.pep                                KVVQFVEXPLRAVVPADSFEPTAQKLNLFK
      orf6a      QIVEHAVLHTPPSFSQSARVVVLFGEEDHKVVQFVEDALRAVVPADSFEPTAQKLNLFK
                  40          50          60          70          80          90

10     orf6.pep      AGAATILFYEDQNVVKGLQEQQPAYAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
      orf6a      AGAATILFYEDQNVVKGLQEQQPAYAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
                  100         110         120         130         140         150

15     orf6.pep      NPLPDAAIKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFEGAX
      orf6a      NPLPDAAIKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFEGAX
                  160         170         180         190         200

```

	1	ATGACCCGTC	AATCTCTGCA	ACAGGCTGCC	GAAAGCCGCC	GTTCCATTTA
20	51	TTCGTTAAAT	AAAAATCTGC	CCGTCGGCAA	AGATGAAATC	GTCCAAATCG
	101	TCGAACACGC	CGTTTTGCAC	ACACCTTCTT	CGTTCAATTG	CCAATCTGCC
	151	CGTGTGGTCG	TGCTGTTTGG	CGAAGAGCAT	GATAAGTGTG	GGCAATTGTG
	201	CGAAGACGCG	TCGCGTGCCG	TCGTGCCTGC	GACACGTTTT	GAACCGACCG
25	251	CGCAAAAATT	GAACCTGTTT	AAGGCGGGTG	CGGCAACTAT	TTTGTTTTAT
	301	GAAGATCAAA	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATGC
	351	CGCCAACTTT	CCCCTTTGGG	CGGACCAGGC	GAACGCGATG	GTCGAGTATG
	401	CCGTTTGGAC	GACACTTGCC	GCGGTCGGCG	TAGGTGCAAA	CCTGCAACAT
30	451	TACAATCCCT	TGCCCCGATG	GGCGATTGCC	AAAGCGTGGA	ATATCCCCGA
	501	AAACTGGTTG	TTGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGGCGG
	551	CAGGTGAAAA	GACCTTTTGA	CCAGTTGCAG	AACGTTTGAA	AGTGTTCGGC
	601	GCATAA				

35

1	MTRQSLQQA	ESRRSIYSLN	KNLPVGKDEI	VQIVEHAVLH	TPSSFSNSQA
51	RVVVLFGEEH	DKVWQFVEDA	LRAVVPADSF	EPTAQKLNLF	KAGAATILFY
101	EDQNVVKGLQ	EQFPAYAAAF	PVWADQANAM	VQYAVVTTLA	AVGVGANLQH
151	YNPLPDAAIA	KAWNIPENWL	LRAQMVGIGI	EGAAGEKTFE	PVAERLKVFG
201	A*				

		50	60	70	80	90	100
40	orf6a.pep	TPSSFNSQSARVVVLFGE	EHDKVWQFVEDALRA	VVPADSF	EP	TAQKLNLFKAGA	ATILFY
	orf6-1					LRAVVPADSF	EP
					10	20	30
45	orf6a.pep	EDQNVVKGLQE	QFPAYAA	NFPVWADQ	ANAMVQYAV	WTTLAAV	GVGANLQHY
	orf6-1	EDQNVVKGLQE	QFPAYAA	NFPVWADQ	ANAMVQYAV	WTTLAAV	GVGANLQHY
50		40	50	60	70	80	90
	orf6a.pep	KAWNIPENWLLRA	QMVI	GGIEGA	AGEKT	FEPVAERL	KLVFGAX
	orf6-1	KAWNIPENWLLRA	QMVI	GGIEGA	AGEKT	FEPVAERL	KLVFGAX
55		100	110	120	130		

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

```

      orf6.pep                                KVVQFVEXPLRAVVPADSFEPTAQKLNLFK    30
      orf6ng      SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVVQFVEDALRAVVPADSFEPTAQKLFK    64
5      orf6.pep      AGAATILFYEDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY    90
      orf6ng      AGAATILFYEDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY    124
10     orf6.pep      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGA    140
      orf6ng      NPLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFFPVAERLKVFGA    174

```

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

```

15      1  ATGGCCGTTG  CGTCAAATGT  CAGCTTGGAT  ATGTCCAATC  CTACGGTGTG
      51  ACGCATGGGA  TTACCCCTTAT  ATATTGCGTC  CCTAAGAAGG  GGCGCAATAT
      101  ATAAGGTGTG  GCAATTTGTC  GAAGACGCGC  TCGTGCCGT  CGTGCCTGCC
      151  GACAGTTTGT  AACCGACCGC  GCAAAAATTG  AAGCTGTGTTA  AGGCGGGCGC
      201  GGCAACCATT  TTGTTTTATG  AAGATCAAAA  TGTCGTCAAA  GGTTCGAGG
      251  AGCAGTTCCC  TGCTTATGCC  GCCAACTTTC  CCGTTTGGGC  GGACCAGGCG
20     301  AACGCTATGG  TACAGTATGC  CGTCTGGACG  ACACTTGCCG  CGGTCGGTGC
      351  AGGTGCAAT  CTGCAACATT  ACAACCCCTT  GCCCGATGTG  GCGATTGCTA
      401  AAGCGTGGAA  TATCCCGAA  AACTGGCTGT  TCGCGCGCA  AATGGTTATC
      451  GGTGGTATTG  AAGGGGcggc  aggtgaaaaa  gtctttgaac  CCGTTGcgga
      501  acgtttgAAA  GTGTTGCGCG  CATAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

```

      1  MAVASNVSLD  MSNPTVLRMG  LPLYIASLRR  GAIYKVVQFV  EDALRAVVP
      51  DSFEPTAQL  KLFKAGAATI  LFYEDQNVVK  GLQEQQPAYA  ANFPVWADQA
      101  NAMVQYAVWT  TLAAVGAGAN  LQHYNPLPDV  AIAKAWNIFE  NWLLRAQPMVI
      151  GGIEGAAGEK  VFEPVAERLK  VFGA*
30

```

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

```

      orf6-1.pep                                10      20      30
      orf6ng      PTVLRMGLPLYIASLRRGAIYKVVQFVEDALRAVVPADSFEPTAQKLNLFKAGAATILFY
      35      20      30      40      50      60      70
      orf6-1.pep      EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDAAIA
      40      40      50      60      70      80      90
      orf6ng      EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDVAIA
      80      90      100      110      120      130
45     orf6-1.pep      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
      orf6ng      KAWNIPENWLLRAQMVIIGGIEGAAGEKVFFPVAERLKVFGAX
      140      150      160      170

```

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

```

55      1  .GGCTACAAC  ACCTGTTCGC  GCGCGGCAGC  CGCATCGCCA  ACTACCAAAT
      51  CAACGGCATC  CCCGTTGCCG  ACGCGCTGGC  CGATACGGG  CAATGCCAAC
      101  ACCGCCGCT  ATGAGCGCGT  AGAAGTCGTG  CGCGGCGTGG  CGGGGCTGCT
      151  GGACGGCAGC  GCGGAGCCTT  CCGCCACCGT  CAATCTGGTG  CGCAAACGCC
      201  TGACCCGCAA  GCCATTGTTT  GAAGTCCGCG  CCGAAGCGGG  CAACCGcAAA

```



-379-

5  
251 CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crC  
301 rCTGCGCgGC CGCCTGGTTT CCAcCTTCGG ACGCGGCGAC TCGTGGCGGC  
351 GGCGCGAAGC CAGCCGskAT GCCGAActCT ACGGCATTTT GGAATACGAC  
401 ATCGCACCGC AAACCCGCGT CCACGCArGC ATGGACTACC AGCAGGCGAA  
451 AGAAACCGCC GACGCGCCGC TCAGcTACGC CGTGTACGAC AGCCAAGGTT  
501 ATGCCACCGC CTTGCGCCCG AAAGACAACC CCGCCACAAA TTGGGCGAAC  
551 AGCCACCACC GTGCGCTCAA CCTGTTCCGC GGCATCGAAC ACCGCTTCAA  
601 CCAAGACTGG AAACCTCAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10  
1 ..GYNLFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL  
51 DGTGEPsATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX  
101 LRGRlVSTFG RGDsWRRER SRXAElyGIL EYDIAPQTRV HAXMDYQOAK  
151 ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSHHRALN LFAGIEHFRN  
201 QDWKLKAEYD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

20  
1 ATGACACGCT TCAAAATATTC CCTGCTGTTT GCGGCCCTGT TGCCCGTGTA  
51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCAAACCG CAGGAAAGCA  
101 CTGAATTGCC GACCATCACC GTTACCGCGC ACCGCACCGC GAGTTC AAC  
151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC  
201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCCTGTTGCA GCGGACCGGC  
301 ACCAGCCGCC AGATTTACGG CTCGACCGC GCGGGCTACA ACTACCTGTT  
351 CGCGCGCGGC AGCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG  
401 CCGACGCGCT GGCGATACG GGCAATGCCA ACACCGCGC CTATGAGCGC  
25  
451 GTAGAAGTCG TGC GCGCGCTG GCGGGGCTG CTGGACGGCA CGGGCGAGCC  
501 TTCCGCCACC GTCAATCTGG TGC GCAAACG CCTGACCGC AAGCCATTGT  
551 TTGAAGTCG CGCGAAGCG GGCAACCGCA AACATTTCCG GCTGGACGGC  
601 GACGTATCGG GCAGCCTGAA CACCGAAGC ACGCTGCGG GCCGCTGCT  
651 TTCCACCTTC GGACGCGCGG ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG  
30  
701 ATGCCGAAct CTACGGCATT TTGGAATACG ACATCGCACC GCAAAACCCG  
751 GTCCACGCAG GCATGGACTA CCAGCAGCGG AAAGAAACCG CCGACGCGCC  
801 GCTCAGCTAC GCCGTGTAG ACAGCCAAGG TTATGCCACC GCCTTCGGCC  
851 CGAAAGACAA CCGCGCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC  
901 AACCTGTTCG CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA  
35  
951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG  
1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC  
1051 GGTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCATTGAT  
1101 CGGCAAATAC CGCTGTTCG GCGCGAACA CGATTTAATC GCGGGTATCA  
1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC  
40  
1201 AACGCCATT CCAACGCCCTA CGAATTTTCC CGCACGGGTG CCTACCGCA  
1251 GCCTGCATCG TTTGCCAAA CCATCCCGCA ATACGGCACC AGGCGGCAAA  
1301 TCGCGGGCTA TCTCGCCACC CGTTTCCGG CCGCCGACAA CCTTTCGCTG  
1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
1401 CACACAAGCG ATGACCTATG TGTCCGCCAA CCGTTTACC CCTACACAG  
45  
1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CTCTTACGG CTCGTACAGC  
1501 AGCCTGTTCG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA  
1551 ACCCGTAACC GGCAACAATC TGAAGCCGG CATCAAAGC GAATGGCTTG  
1601 AAGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCG TAAAAACAAC  
50  
1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC  
1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCGCGCA  
1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCG  
1801 GACCAAGACG GCAGCCGCTT GAACCCCGAC AGCGTACCG AACGCAGCTT  
1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCC AGCGGCTGGA  
55  
1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC  
1951 ACGCTCCGCA TCCCAACCC CGCCGCCAAA GCGCGCGCG CCGACAACAG  
2001 CCGCAAAAA GCCTACGCCG TCGCGACAT CATGGCGCTG TACCGCTCA  
2051 ATCCGCGCGC CGAACTGTG CTGAACGTGG ACAATCTGTT CAACAAACAC  
2101 TACCGCACCC AGCCGACCG CCACAGCTAC GCGCAGTGC GGACAGTGAA  
2151 CGCGCGGTTT ACCTATCGGT TTAATAA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

65  
1 MTRFKYSLLE AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN  
51 DGYTVSGTHT PLGLEMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG  
101 TSRQIYGS DR AGYNLFAR G RIANYQING IPVADALADT GNANTAAYER  
151 VEVVRGVAGL LDGTGEPsAT VNLVRKRLTR KPLFEVRAEA GNRKHFGLD A  
201 DVSGSLNTEG TLRGRlVSTF GRGDsWRRR RSRDAELYGI LEYDIAPQTR

-380-

5  
10  
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL  
301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HNTAATDLIP  
351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP  
401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL  
451 ILGGRYTRYR TGSYDSRTQG MTYVSANRET PYTGIVFDLT GNLSLYGSYS  
501 SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRNAS AAVYRARKNN  
551 LATAAGRDP S GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR  
601 DQDGSRLNPD SVFERSEKLF TAYHFAPEAP SGWTIGAGVR WQSETHDPA  
651 TLRI PNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNDL FNKH  
701 YRTQPD RSY GALRTVNAAF TYRFK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

15 Orf23 6 FARGSRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEPSATVNLVRK 65  
++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK  
PupB 215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273  
20 Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFXXXXXXXXXXXXXAE 125  
R T + + EAGN +G DVSG L +RGR V+ +  
PupB 274 RPTAEAQASITGEAGNWDRYGTGFVDVSGPLTETGNIRGRFVADYKTEKAWIDRYNQSQL 333  
25 Orf23 126 LYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183  
+YGI E+D++ T + Y + D+PL + S G T N A +W+  
PupB 334 MYGITEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRSLNAAPDWSY 391  
30 Orf23 184 SHHRALNLFAGIEHRFNQDWKLKAE 208  
+ H + F IE + W K E  
PupB 392 NDHEQTSFFTSIEQQLGNGWSGKIE 416

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N. meningitidis*:

35 orf23.pep 10 20 30  
GYNLYFARGSRIANYQINGIPVADALADTG  
|||  
orf23a QMRDQNIKALDRALLQATGTSRQIYGSDRAGYNLYFARGSRIANYQINGIPVADALADTG  
90 100 110 120 130 140  
40 orf23.pep 40 50 60 70 80 90  
NANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDAD  
|||  
orf23a NANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRPTRKPLFEVRAEAGNRKHFGLGAD  
150 160 170 180 190 200  
45 orf23.pep 100 110 120 130 140 150  
VSGSLNTEXXLRGRLVSTFGRGDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQA  
|||  
orf23a VSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGILEYDIAPQTRVHAGMDYQQA  
210 220 230 240 250 260  
50 orf23.pep 160 170 180 190 200 210  
ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRALNLFAGIEHRFNQDWKLKAEYD  
|||  
orf23a ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRALNLFAGIEHRFNQDWKLKAEYD  
270 280 290 300 310 320  
55 orf23.pep Y  
I  
60 orf23a YTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIA  
330 340 350 360 370 380

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

-381-

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCCGTGA  
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCAAAACCG CAGGAAAGCA  
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC  
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC  
 5 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
 251 GCGACCAAAA CATCAAAGCG CTCGACCGCG CCCTGTTGCA GCGCACCAGC  
 301 ACCAGCCGCC AGATTTACGG CTCGACCGCG GCGGGCTACA ACTACCTGTT  
 351 CGCGCGCGGC AGCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG  
 401 CCGACGCGCT GGCGGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC  
 10 451 GTAGAAGTCG TGCGCGGCGT GCGGGGCTG CTGGACGGCA CGGGCGAGCC  
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCCGACCGCG AAGCCATTGT  
 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCCG GCTGGCGCGC  
 601 GACGTATCGG GCAGCCTGAA TGCCGAAGGC ACGCTGCGCG GCCGCCTGCT  
 651 TTCCACCTTC GGACGCGCGC ACTCGTGGCG GCAGCGCGAA CGCAGCCGCG  
 15 701 ATGCCGAACCT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCGC  
 751 TTCCACGCGAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC  
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC  
 851 CGAAAGACAA CCCC GCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC  
 901 AACCTGTTTC CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA  
 20 951 AGCCGAATAC GACTACACCC GACGCGGCTT CCGCCAGCCC TACGCGCTAG  
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC  
 1051 GGTATTATGCG ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCATTAAAT  
 1101 CGGCAATAC GCCTGTGTCG GCCGCGAACA CGATTTAATC GCGGGTATCA  
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCGA CATCATCCCC  
 25 1201 AACGCCATTC CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA  
 1251 GCCTGCATCG TTTGCCCAAA CCATCCCGCA ATACGGCACC AGGCGGCAAA  
 1301 TCGCGCGCTA TCTGCCACC CGTTTCCGCG CCGCCGACAA CTTTTCGTG  
 1351 ATACTCGGCG GCAGATACAG CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACCC CCTACACAG  
 30 1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC  
 1501 AGCCTGTTTC GACGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGAAGCCGG CATCAAAGGC GAATGGCTTG  
 1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAAC  
 35 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC  
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCGCCGA  
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC  
 1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCCG AACGCAGCTT  
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA  
 40 1901 CCATCGGCGC AGGCGTGCAG TGGCAGAGCG AAACCCACAC CGACCTGCC  
 1951 ACGCTCCGCA TCCCCAACCC CGCCGCCAAA GCGCGCGCCG CCGACAACAG  
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCGC CGAACTGTG CTGAACGTGG ACAATCTGTT CAACAAACAC  
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCAGTGC GGACAGTGAA  
 2151 CGCGGCGTTT ACCTATGGGT TTAAATAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLE AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT FLGLEMTLRE IPQSVSVITS QMRDQNIKA LDRALLQATG  
 101 TSRQIYGS DR AGYNLFARG SRIANYQING IPVADALADT GNANTAAYER  
 151 VEVVRGVAGL LDGTGEPSAT VNLVRKRPT R KPLFEVRAEA GNRKHFGLGA  
 50 201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR  
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL  
 301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HNTAATDLIP  
 351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP  
 401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL  
 55 451 ILGGYRSYR TGSYDSRTQG MTYVSANRET PYTGIVFDLT GNLSLYGSYS  
 501 SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRNLAS AAVYRARKNN  
 551 LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR  
 601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSEHTDPA  
 651 TLRI PNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH  
 60 701 YRTQDRHSY GALRTVNAAF TYRFK\*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

orf23a.pep 10 20 30 40 50 60  
 MTRFKYSLLEAALLPVYAQADVSVSDPPKPQESTELPTITVTADRTASSNDGYTVSGTHT  
 65 orf23-1 10 20 30 40 50 60  
 MTRFKYSLLEAALLPVYAQADVSVSDPPKPQESTELPTITVTADRTASSNDGYTVSGTHT

-382-

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGS DRAGYNL FARG					
5	orf23-1	PLGLPMTLREIPQSVSVITSQQMRDQNIKTLD RALLQATGTSRQIYGS DRAGYNL FARG					
		70	80	90	100	110	120
	orf23a.pep	130	140	150	160	170	180
10	orf23-1	SRIANYQINGIPVADALADTGNANTAA YERVEVVRGVAGLLDGTGEPSATVNLVRKR PTR					
		130	140	150	160	170	180
	orf23a.pep	190	200	210	220	230	240
15	orf23-1	KPLFEVRAEAGNRKHFGLGADVSGSLNAE GTLRGRLVSTFGRGDSWRQRERSRDAELYGI					
		190	200	210	220	230	240
	orf23a.pep	250	260	270	280	290	300
20	orf23-1	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL					
		250	260	270	280	290	300
25	orf23a.pep	310	320	330	340	350	360
	orf23-1	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH					
30		310	320	330	340	350	360
	orf23a.pep	370	380	390	400	410	420
	orf23-1	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
35		370	380	390	400	410	420
	orf23a.pep	430	440	450	460	470	480
40	orf23-1	FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
		430	440	450	460	470	480
	orf23a.pep	490	500	510	520	530	540
45	orf23-1	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRLNAS					
		490	500	510	520	530	540
50	orf23a.pep	550	560	570	580	590	600
	orf23-1	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
		550	560	570	580	590	600
55	orf23a.pep	610	620	630	640	650	660
	orf23-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSETHD PATLRIPNPAAK					
60		610	620	630	640	650	660
	orf23a.pep	670	680	690	700	710	720
	orf23-1	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPPDRHSYGALRTVNAAF					
65		670	680	690	700	710	720
	orf23a.pep	TYRFXK					
70	orf23-1	TYRFXK					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N. gonorrhoeae*:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNYLFARGSR IANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLPD	60
10	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLVSTFGR	111
	orf23ng	GTGEPSATVNLVRKHPTKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLVSTFGR	120
15	orf23.pep	GDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWQLERSRDAELYGILEYDIAPQTRVHAGMDYQAKETADAPLSYAVYDSQGYATAF	180
	orf23.pep	GPKDNPATNWANSRHHRALNLFAGIEHREFNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNRNRALNLFAGIEHREFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising

20 amino acid sequence <SEQ ID 670>:

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEPSATVN	LVRKHPTKRP	LFEVRAEAGN	RKHFGGLGADV
	101	SGSLNAEGTL	RGRLVSTFGR	GDSWQLERS	RDAELYGILE	YDIAPQTRVH
	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSNRNALNL
25	201	FAGIEHREFN	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
	301	IPNAYEFSRT	GAYPQPSSFA	QTIPQYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGYRSRYRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
	401	FVPQLQKDEH	GSYLKPVGTN	NLEADIKGEW	LEGRLNASAA	VYRARKNNLA
30	451	TAAGRQDSGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQA	GYSQSKPRDQ
	501	DGSRLNPDSV	PERSFKLFET	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTLSLN	VDNLFNKHRY
	601	TQPDHRSYGA	LRTVNAAFY	REF*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CGGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCGC	GAGTTCACAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCCGGC	TGCCCATGAC
	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
40	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAACCTACCA	AATCAACGGC	ATCCCCGTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGGACGGCA	CGGGCGAGCC
45	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAAACCCG
50	751	GTCCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCGGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCTG	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCG	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
55	1001	CAGGCGTACT	TTCCATGTGC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCcgatcc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC
	1101	CGGCAATATC	CgcctGTTCG	GCCGCGAGCA	CGATTTAATC	GCGGGTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCGC	CATCATTTCC
	1201	AACGCCATT	CCAACGCCCTA	CGAATTTTCC	CGCACGGGCG	CCTATCCGCA
60	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTCGCTG
	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCGC	GCAGGACGCT	ACAACAGCCG

-384-

5  
10  
15

```

1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG
1451 GCATCGTGTG CGATCTGACC GGCAACCTGT CGTTTACGG CTCGTACAGC
1501 AGCCTGTTTCG TCCCGCAATT GCAAAAAGAC GAACACGGCA GCTACCTGAA
1551 ACCCGTAACC GGCAACAATC TGAAGCCGA CATCAAAGGC GAATGGCTTG
1601 AAGGGCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
1651 CTCGCCACCG CAGCAGGACG CGACCAGAGC GGCAACACCT ACTATCGCGC
1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCGCCGA
1751 TCACGCCCGA ATGGCAGATA CAGGCAGGCT ACAGCCAAAG CAAACCCCGC
1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTAcCG AACGCAGCTT
1851 CAAACTCTTC ACCGCCTACC ACTTAGCCCC CGAAGCCCCC AGCGGCCGGA
1901 CCATcggTGC GGGTGTGCGC CGGCAGGGCG AAACCCACAC CGACCCAGCC
1951 GCGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG TCGCCAACAG
2001 CCGCCAGAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
2051 ATCCGCGCAC CGAACTGTCT CTGAACGTGG ACAACCTGTT CAACAAACAC
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCGACTGC GGACAGTGAA
2151 CGCGGCGTTT ACCTATCGGT TTAATAAA

```

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

20  
25  
30

```

1 MTRFKYSLLF AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN
51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG
101 TSRQIYGS DR AGYNLFARG SRIANYQING IPVADALADT GNANTAAYER
151 VEVVRGVAGL PDGTGEP SAT VNLVRKHPT R KPLFEVRAEA GNRKHFGLGA
201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQLE RSRDAELYGI LEYDIAPQTR
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNSRNRAL
301 NLFAGIEHRE NQDWKLA EY DYTRSRFRQ PYGVAGVLSID HSTAATDLIP
351 GYWHADPRTH SASMSLTGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
401 NAIPNAYEFS RTGAYPOPSS FAQTIPOYDT RRQIGGYLAT RFRAADNLSL
451 ILGGYRSRYR AGSYNSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
501 SLFVPQLQKD EHGSYLKPV T GNNLEADIKG EWLEGRNLAS AAVYRARKNN
551 LATAAGRDQS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKPR
601 DQDGSRLNPD SVPERSFKLF TAYHLAPEAP SGRITIGAGVR RQGETHTDPA
651 ALRIPNPAK ARAVANSRQK AYAVADIMAR YRFNPRTELS LNVNDFNKH
701 YRTQDRHSY GALRTVNAAF TYRFK*

```

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

35  
40  
45  
50  
55  
60  
65

```

              10      20      30      40      50      60
orf23-1.pep  MTRFKYSLLF AALLPVYAQADVSVSDPPKPQESTELPTITVTADRTASSNDGYTVSGTHT
orf23ng-1    MTRFKYSLLF AALLPVYAQADVSVSDPPKPQESTELPTITVTADRTASSNDGYTVSGTHT
              10      20      30      40      50      60

              70      80      90      100     110     120
orf23-1.pep  PLGLPMTLREIPQSVSVITSQQMRDQNIKTLD RALLQATGTSRQIYGS DRAGYNLFARG
orf23ng-1    PFGLPMTLREIPQSVSVITSQQMRDQNIKTLD RALLQATGTSRQIYGS DRAGYNLFARG
              70      80      90      100     110     120

              130     140     150     160     170     180
orf23-1.pep  SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTR
orf23ng-1    SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPDGTGEP SATVNLVRKHPT R
              130     140     150     160     170     180

              190     200     210     220     230     240
orf23-1.pep  KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRVSTFGRGDSWRRRERSRDAELYGI
orf23ng-1    KPLFEVRAEAGNRKHFGLCADVSGSLNAEGTLRGRVSTFGRGDSWRQLERSRDAELYGI
              190     200     210     220     230     240

              250     260     270     280     290     300
orf23-1.pep  LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNSRHRAL
orf23ng-1    LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNSRNRAL
              250     260     270     280     290     300

              310     320     330     340     350     360
orf23-1.pep  NLFAGIEHREFNQDWKLA EYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH
orf23ng-1    NLFAGIEHREFNQDWKLA EYDYTRSRFRQPYGVAGVLSIDHSTAATDLIPGYWHADPRTH

```

-385-

		310	320	330	340	350	360
5	orf23-1.pep	370	380	390	400	410	420
	orf23ng-1	370	380	390	400	410	420
10	orf23-1.pep	430	440	450	460	470	480
	orf23ng-1	430	440	450	460	470	480
15	orf23-1.pep	490	500	510	520	530	540
	orf23ng-1	490	500	510	520	530	540
20	orf23-1.pep	550	560	570	580	590	600
	orf23ng-1	550	560	570	580	590	600
25	orf23-1.pep	610	620	630	640	650	660
	orf23ng-1	610	620	630	640	650	660
30	orf23-1.pep	670	680	690	700	710	720
	orf23ng-1	670	680	690	700	710	720
35	orf23-1.pep	TYRFXK					
	orf23ng-1	TYRFXK					

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
50	>gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotruclic acid precursor [Escherichia coli] Length = 729 Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPOSVSVITSQQMRDQNIKTLDRL 95 T+ V TA + + Y+V+ T + MT R+IPQSV+++ Q+M DQ ++TL + Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTSAGTKMQMTQORDIPQSVTIVSQRMEDQQQLQTLGEVM 102
60	Query: 96 LQATGTSRQIYGSDRAGYNLFARGSRANYQINGIP-----VADALADTGNANTAA 147 G S+ SDRA Y ++RG +I NY ++GIP + DAL+D A Sbjct: 103 ENTLGISKSQADSDRALY---YSRGFQIDNYMVDGIPTYFESRWNLGDALSDM-----AL 154
65	Query: 148 YERVEVVRGVAGLPDGTGEP SATVNLVRKHPTKPLF-EVRAEAGNRKHFGLGADVSGSL 206 +ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSPL 214
70	Query: 207 NAEGLRGRILVSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAKETADA 266 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGGYQNDSWLDRYNSEKTFFGIVDADLDLTLTSLAGYEYQRIDVNSPT 274
	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNSRNALNLFAGIEHRFNQDWKLAKEYDYTRSR 326

-386-

```

          +++ G + ++      + A +W+ +      +F ++ +F  W+      ++
Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334

5  Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADFRTHSA-SMSLTGKYRLFG 374
      F + Y A V D      ++ PG+      W++ R A + G Y LFG
Sbjct: 335 FDSKMMYVDAYVNKADGMLVGPFYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSEYELFG 394

10 Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYFPQSSFAQTIPQYDTRR 432
      R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT
Sbjct: 395 RQHNLMFQ-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451

15 Query: 433 QIGGYLATRFRAADNLSLILGGYSRYRAGSYNSRTQGMTY-VSANRFTPYTGIVFDXXX 491
      Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD
Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFIND 504

20 Query: 492 XXXXXXXXXXXXFFVQLQKDEHGSYLKPVGTGNNLEADIKGEWLEGRINASAAYVRARKNNL 551
      F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
Sbjct: 505 NWSTYASYTSIFQPQNDRDSSGKYLAPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564

25 Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKPRDQDGSRLN 608
      A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVEFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624

30 Query: 609 PDSVPERSEFKLFTAYHLAEPASGRTIGAGVRRQGETHTDPAALRIPNPAKARAVANSR 668
      P ++P + K+FT+Y L P P T+G GV Q +TD P RA
Sbjct: 625 P-NLPRTTVKMFYSYRL-PVMPE-LTVGGGVNWQNRVYTDTV-----TPYGTFR-----E 672

Query: 669 QKAYAVADIMARYRFNPRTELSLNVNDLNFNKHRYTQPDRH-SYGALRTVNAAFTYRF 724
      Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F
Sbjct: 673 QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTPRNFSITGTYQF 729

```

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

#### 40 Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

```

1  ATGCGCACGG CAGTGGTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
45 151 AGCGTCAGcA CGCCTGCTTC GCGGgcGgCa ATCATACCTT CGTCTTCGGA
201 AACGGGGATA AACGcGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAGCCCAT TnTTCAAGAA TGCGTGCCAC
351 TnAGTCGCCG ACGGGG..

```

50 This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
51  SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
101 PCVPQTLKPI XSRMRATXSP TG..

```



Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

      1  ATGCGCACGG  CAGTGGTTTT  GCTGTTGATC  ATGCCGATGG  CGGCTTCGTC
      51  GGCAATGATG  CCGGAAATGG  TGTGCGCGGG  CGTGTGCGCG  GGAACGGCAA
    101  TCATATCCAA  GCCGACCGAA  CAAACGGCGG  TCATGGCTTC  GAGTTTGTCC
    151  AGCGTCAGCA  CGCCTGCTTC  GCGGCGGGCA  ATCATACCTT  CGTCTTCGGA
    201  AACGGGGATA  AACGCGCCAC  TCAAACCCCC  GACCGCGCTG  GAAGCCATCA
    251  TGCGGCCTTT  TTTCACGGCA  TCGTTCAGCA  ATGCCAAAGC  TGCTGTTGTG
    301  CCGTGCGTAC  CGCAGACGCT  CAAGCCCAT  TCTTCAAGAA  TGCGTGCCAC
    351  TGAGTCGCGG  ACGGCGGGGG  TCGGCGCCAG  CGACAAGTCG  AGAATACCAA
    401  ACGGGATATT  CAGCATTTTT  GAGGCTTCGC  GGCCGATGAG  TTCGCCACG
    451  CGGGTAATTT  TGAAAGCAGT  TTTCTTCACT  ACTTCCGCAA  CTTCGGTCAG
    501  TGTCGTTGCA  TCTGAATTTT  CCAACGCGGC  TTTTACGACA  CCTGGGCGCG
    551  ATACGCCGAC  ATTGATAACG  GCATCCGCTT  CGCCCGAACC  ATGAAACGCG
    601  CCCGCCATAA  ACGGGTTGTC  TTCCACCGCG  TTGCAGAAAC  CGACAATTTT
    651  AGCGCAGCCG  AAACCTTCGG  GCGTGATTTC  CGCCGTGCGT  TTGACGGTTT
    701  CGCCGCGCAG  CTTGACCGCA  TCCATATTGA  TACCGGCACG  CGTACTGCCG
    751  ATATTGATGG  AGCTGCACAC  AATATCGGTA  GTCTTCATCG  CTTCGGGAAT
    801  GGAGCGGATT  AACACCTCAT  CCGAAGGCGA  CATCCCTTTT  TGCACCAACG
    851  CGGAAAAACC  GCCGATAAAA  GACACACCGA  TGGCTTTGGC  AGCTTTATCC
    901  AAAGTTTGCG  CCACGCTGAC  GTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

      1  MRTAVVLLLI  MPMAASSAMM  PEMVCAGVSP  GTAIISKPTE  QTAVMASSLS
      51  SVSTPASAAA  IIPSSSETGI  NAPLKPPTAL  EAIMPPFFTA  SFSNAKAAVV
    101  PCVPQTLKPI  SSRMRATESP  TAGVGASDKS  RIPNGIFSIF  EASRPMSSPT
    151  RVILKAVFFT  TSATSVNVVA  SEFSNAAFTT  PGPDTPTLIT  ASASPEP*NA
    201  PAINGLSSTA  LQNTTILAQP  KPSGVISAVR  LTVSPASLTA  SILIPARVLP
    251  ILMELHTISV  VFIASGMERI  NTSSEGDIPF  CTNAEKPPIK  DTPMALAALS
    301  KVCATLT*
  
```

Computer analysis of this amino acid sequence gave the following results:

### 30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
35	orf24a.pep	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIISKPTE	QTAVMASSLS	SVSTPASAAA
	orf24	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIISKPTE	QTAVMASSLS	SVSTPASAAA
		10	20	30	40	50	60
40	orf24a.pep	IIPSSSXT	GINAPLKPTALE	AIMPPFFTA	SFSNAKAAVV	PCVPQTLKPI	SSRMRATESP
	orf24	IIPSSSET	GINAPLKPTALE	AIMPPFFTA	SFSNAKAAVV	PCVPQTLKPI	SSRMRATESP
		70	80	90	100	110	120
45	orf24a.pep	TAGVGASDKS	RIPNGIFSIF	EASRPMSSPT	RVILKAVFFT	TSATSVNVVA	SEFSNAAFTT
	orf24	TAGVGASDKS	RIPNGIFSIF	EASRPMSSPT	RVILKAVFFT	TSATSVNVVA	SEFSNAAFTT
		130	140	150	160	170	180
50	orf24a.pep	PGPDTPTLIT	ASASPEPXNAP	AIKGLSSXAL	QNTTILAQP	KPSSVISXVRL	MLMVSPASLTA
	orf24	PGPDTPTLIT	ASASPEPXNAP	AIKGLSSXAL	QNTTILAQP	KPSSVISXVRL	MLMVSPASLTA
		190	200	210	220	230	240
55	orf24a.pep	SILIPARVLP	ILMELHTISV	VFIASGMER	NTSSEGDIPF	CTSAEKPPIK	DTPMALAALS
	orf24	SILIPARVLP	ILMELHTISV	VFIASGMER	INTSSEGDIPF	CTNAEKPPIK	DTPMALAALS
60		250	260	270	280	290	300

-388-

```
orf24a.pep  KVCATLTX
            |||||
orf24       KVCATLTX
```

5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

```
1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG TGTGTCGCCG GGAACGGCAA
101 TCATATCCAA NCCGACCGAA CAAACGGCGG TCATCGCTTC GAGTTTATCC
151 AACGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
201 NACGGGGATA AACGCGCCAC TCAAACCGCC AACCGCGCTC GAAGCCATCA
251 TGCCGCCCTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTGTGTG
301 CCGTGCGTAC CGCAGACGCT CAAACCCATT TCTTCAAGAA TGCGCGCCAC
351 CGAGTCGCGG ACGGCAGGGG TCGGTGCCAG CGACAAGTCG AGAATACCAA
401 ACGGGGATAT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TCGCCACAG
15 451 CGGGTAATTT TGAAGGCGGT TTTCTTCACA ACTTCGGCAA CTTCGGTCAA
501 TGTGCTTGCA TCCGAATTTT CCAACGCGGC TTTTACGACA CCCGGGCGCG
551 ATACGCCGAC ATTAATCACA GCATCCGCTT CGCCTGAGCC GTGAAACGCG
601 CCGGCCATAN ACGGGTTGTC TCCNCCGCG TTGCAGAACG CACGATTTT
20 651 GGCGCAGCCG AAACCTTCTA GTGTGATTTC ANCCGTGCGT TTGATGGTTT
701 CGCCCGCCAG TCTGACCGCG TCCATATTGA TACCGGCGCG CGTACTGCCG
751 ATATTGATGG AGCTGCACAC GATATCAGTA GTCTTCATCG CTTCGGGAAT
801 GGAACGGATN AACACCTCGT CAGAAGGCGA CATACTTTT TGCACGACG
851 CGGAAAAGCC GCCAATAAAA GACACGCCGA TGGCTTTGGC AGCCTTATCC
901 AAAGTTTGCG CCACGCTGAC GTAA
```

25 This encodes a protein having amino acid sequence <SEQ ID 678>:

```
1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS
51  NVSTPASAAA IIPSSSXTGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAV
101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT
151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDTPTLIT ASASPEP*NA
201 PAIXGLSSXA LQNTTILAQP KPSSVISXVR LMVSPASLTA SILIPARVLP
251 ILMELHTISV VFASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
301 KVCATLT*
```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

```
35      10      20      30      40      50      60
orf24a.pep  MRTAVVLLLIMPMAASSAMMPMV CAGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
            |||||
orf24-1     MRTAVVLLLIMPMAASSAMMPMV CAGVSPGTAIISKPTQTAVMASSLSNVSTPASAAA
            |||||
40      10      20      30      40      50      60
orf24a.pep  IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAAVPCVPQTLKPISSRMRATESP
            |||||
orf24-1     IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAAVPCVPQTLKPISSRMRATESP
            |||||
45      70      80      90      100     110     120
orf24a.pep  TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
            |||||
orf24-1     TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
            |||||
50      130     140     150     160     170     180
orf24a.pep  PGPDTPTLITASASPEPXNAPAI XGLSSXALQNTTILAQKPSSVISXVRLMVSPASLTA
            |||||
orf24-1     PGPDTPTLITASASPEPXNAPAI NGLSSXALQNTTILAQKPSGVISAVRLTVSPASLTA
            |||||
55      190     200     210     220     230     240
orf24a.pep  SILIPARVLPILMELHTISVVF IASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
            |||||
orf24-1     SILIPARVLPILMELHTISVVF IASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS
            |||||
60      250     260     270     280     290     300
```

orf24a.pep KVCATLTX  
 5 orf24-1 KVCATLTX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

10	orf24.pep	MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTQTAVMASSLSVSTPASAAA	60
	orf24ng	MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIMSKPTQTAVMASSLSVNTPASAAA	60
15	orf24.pep	IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPIXSRMRATXSP	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRATESP	120
	orf24.pep	TG	122
20	orf24ng	TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFFTSATSVRLTASEFSSAALT	180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

1	ATGCGCACCG	CGGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
51	GGCGATGATG	CCGCAAATGG	TGTGCGCGGG	CGTGTGCGCCG	GGAACGGCAA
101	TCATGTCCAA	ACCAACGGAG	CAGACGGCGG	TCATGGCTTC	GAGTTTGTCC
25	151	AGCGTCAACA	CGCCTGCCTC	GGCGGCGGCA	ATCATACCTT
	201	AACGGGGATA	AACGCGCCGC	TCAAACCGCC	GACCGCGCTG
	251	TGCCGCCCTT	TTTACAGGCA	TCGTTAGCA	ATGCCAAAGC
	301	CCGTGCGTAC	CGCAGACGCT	CAAGCCCAT	TCTTCAAGAA
	351	CGAGTCGCG	ACGCGGGGG	TCGGTGCCAG	CGACAAATCG
30	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GACCGATGAG
	451	CGGGTGATTT	TGAAAGCGGT	TTTCTTCACG	ACTTCGGCGA
	501	GCTGACCGCG	TCCGAATTTT	CCAGCGCGGC	TTTGACCACG
	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCCGAGCC
	601	CCCGCCATAA	ACGGATTGTC	TTCCACCGCG	TTGCAGAACA
35	651	GGCGCAGCCG	AAACCTTCGG	GTGTGATTTC	AGCCGTGCGT
	701	CGCCTGCCAG	CTTGACCGCA	TCCATATTGA	TACCGGCACG
	751	ATATTGATGG	AGCTGCACAC	GATATCGGTA	GTTTTTCATC
	801	GGAACGGATC	AACACCTCAT	CCGAAGGCGA	CATACCTTTT
	851	CGGAAAAGCC	GCCGATAAAG	GACACGCCGA	TGGCTTTGGC
40	901	AAAGTCTGCG	CCACGCTGAC	ATAA	

This encodes a protein having amino acid sequence <SEQ ID 680>:

1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIMSKPTE	QTAVMASSLS
51	SVNTPASAAA	IIPSSSETGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAAVV
101	PCVPQTLKPI	SSRMRATESP	TAGVGASDKS	RMPNGIFSIF	EASRPMSSPT
45	151	RVILKAVFFT	TSATSVRLTA	SEFSSAALT	PGDPTPLIT
	201	PAINGLSSTA	LQNTTILAQP	KPSGVISAVR	LMVSPASLTA
	251	ILMELHTISV	VFIASGTERI	NTSSEGDIPF	CTSAEKPIIK
	301	KVCATLT*			DTPMALAALS

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

50	orf24-1.pep	10	20	30	40	50	60
	orf24ng	10	20	30	40	50	60
55	orf24-1.pep	70	80	90	100	110	120
60	orf24ng	70	80	90	100	110	120

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRIPNGIFSI FEASRPMSSPTRVILKAVFFTTTSATSVNVVASEFSNAAFTT					
	orf24ng	TAGVGASDKSRMPNGIFSI FEASRPMSSPTRVILKAVFFTTTSATSVRLTASEFSSAALT					
5		130	140	150	160	170	180
	orf24-1.pep	PGPDTPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
	orf24ng	PGPDTPTLITASASPEPWNAPAINGLSSTALQNTTILAQPKPSGVISAVRLMVSPASLTA					
10		190	200	210	220	230	240
	orf24-1.pep	SILIPARVLPILMELHTISVVFIA SGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
	orf24ng	SILIPARVLPILMELHTISVVFIA SGTERINTSSEGDIPFCTSAEKPPIKDTPMALAALS					
15		250	260	270	280	290	300
	orf24-1.pep	KVCATLT					
	orf24ng	KVCATLT					
20		250	260	270	280	290	300

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-  
underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein,  
it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could  
be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

```

1  ..ACCGACGTGC AAAAAGAGTT GGTCGGCGAA CAACGCAAGT GGGCGCAGGA
51  AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGAC CGGCAGGAAT
101 ACGCCGAATA CCTCAAGCTG CAATGCGACA CGCGGATGAC GCGCGAACGG
151 ATACAGTATC TTCGGGCTA TTCCATCGAT TAG

```

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

```

1  ..TDVQKELVGE QRKWAQEKIS NCRQAAQAD RQEYAEYLKL QCDTRMTRER
51  IQYLRGYSID *

```

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

```

1  ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
51  CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
101 TGCAAGGCAT ACGCGGCAAT ATTCAGGAAA CGCTCAGCCA GGAAGCGCGT
151 TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATGCCG ACAAATATAT
201 CGCCGCCGCC TACGGTTTGG CGTTTTCTTT GGAACACGCT TCGGAAACGC
251 AGGAAGGCGG GCGCACGTTC TGTATCGCCG ATTTGAACAT TACCGTGCCG
301 TCTGAAACGC TTGCCGATGC CAAGGCAAAC AGCCCCCTGT TGTACGGGGA
351 AACTGCTTTG TCGGATATTG TCGGCAGAAA GACGGGCGGC AATGTCGAGT
401 TTAAGACGCG CGTATTGACG GCAGCCGTCC GCTTCCTGCC CGTCAAAGAC
451 GGTGAGACGG CATTGTGCGA CAACACGCTC GGTATGGCGG CGCAAACGCT
501 GTCTGCCGCG CTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT GAAAAAGAAA GACGCGGTCA GGATTTTGAG CGGAAAAGCC
601 CGTGAAGAAG AACCGTCCAA ACCACGCCC GAAGACATTT TGGAACACAA
651 TGCCGCCGCG GCGATGCGG GCGTACCCCA AGCCGCAGAA GGCAGCGCCG
701 AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
751 GTATCACGGG GCGAAGTGGA AGAGGCGCGC GTACAAAACC AGCGTGCGGA
801 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
851 AGTTGGTCGG CGAACAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
901 CGACAAGCCG CCGCGCAGGC AGACGGGCGG GAATACGCCG AATACCTCAA
951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
1001 GCTATTCCAT CGATTAG

```

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

      1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEAR
      51 SFAREDXQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
    101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGLT AAVRFLPVKD
    151 GQTAFVDNTV GMAAQTLASA LLPYGVKSIV MIDGKAVKKE DAVRILSGKA
    201 REEPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
    251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEXR KWAQEKISNC
    301 RQAAAQADRQ EYAEYLKLQC DTRMTRERIQ YLRGYSID*
  
```

Computer analysis of this amino acid sequence gave the following results:

#### 10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N. meningitidis*:

```

      10      20      30
    15 orf25.pep      TDVQKELVGEQRKWAQEKISNCRQAAAQAD
      orf25a      VTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRWAQEKISNCRQAAAQAD
      250      260      270      280      290      300

      40      50      60
    20 orf25.pep      RQEYAEYLKLQCDTRMTRERIQYLRGYSIDX
      orf25a      RQEYAEYLKLQCDTRMTRERIQYLRGYSIDX
      310      320      330
  
```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

    25      1 ATGTATCGGA AACTCATTCG GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
      51 CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
    101 TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCACGCA GGAAGCGCGT
    151 TCTTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATGCCG ACNAAATTAT
    201 CGCCGCCGCC TANGNTNNGN NGNTNTCTTT GGAACACGCT TCGGAAACGC
    301 TCTGAAACGC TTGCCGATGC CAAGGCAAAC AGCCCCCTGC TGTACGGGGA
    351 AACCGCTTTG TCGGATATTG TCGGCAGAGG GACGGGCGGC AATGTCGAGT
    401 TTAAAGACGG CGTATTGACG GCAGCCGTCC GCTTCTTACC CGTCAAAGAC
    451 GGTGAGANGG CATTTGTCTG CAACACGGTC GGTATGGCGG CGCAAACGCT
    501 GTCTGCCGCG TTGCTGCCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
    551 GCAAGGCGGT AAAAAAGAA GACGCGGTCA GGATTNTGAG CNGANAAGCC
    601 CGTGAANAAG AACCGTCCAA ANCCNNGCCC GAAGACATT TGGAAACATAA
    651 TGCCGCCGGA GGGGATGCAG ACGTACCCCA AGCGGAGAGG GACGCGCCCG
    701 AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
    40 751 GTATCACGGG GCGAAGTGGA AGAGGCGCGN GTACAAAACC AGCGTCCGGA
    801 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
    851 AGTTGGTCCG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
    901 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
    951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
    45 1001 GCTATTCCAT CGATTAG
  
```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

      1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
      51 SFAREDXQF VDADKIIAAA XXXXSLEHA SETQEGGRTF CXADLNITVP
    101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGLT AAVRFLPVKD
    151 GQXAFVDNTV GMAAQTLASA LLPYGVKSIV MIDGKAVKKE DAVRIXSXXA
    201 REXEPSKXXP EDILEHNAAG GDADVPQAGE DAPEPEILHP DDGERADTVT
    251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEXR KWAQEKISNC
    301 RQAAAQADRQ EYAEYLKLQC DTRMTRERIQ YLRGYSID*
  
```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

    55      10      20      30      40      50      60
      orf25a.pep MYRKLIALPFALLLAACGRE EPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXQF
      orf25-1 MYRKLIALPFALLLAACGRE EPPKALECANPAVLQGIIRXNIQETLTQEARSFAREDXQF
  
```

-392-

		10	20	30	40	50	60
5	orf25a.pep	70	80	90	100	110	120
	orf25-1	70	80	90	100	110	120
10	orf25a.pep	130	140	150	160	170	180
	orf25-1	130	140	150	160	170	180
15	orf25a.pep	190	200	210	220	230	240
	orf25-1	190	200	210	220	230	240
20	orf25a.pep	250	260	270	280	290	300
	orf25-1	250	260	270	280	290	300
25	orf25a.pep	310	320	330	339		
	orf25-1	310	320	330	330		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from  
 35 *N.gonorrhoeae*:

orf25.pep	TDVQKELVGEQRKWAQEKISNCRQAAAQAD	30
orf25ng	VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNCRQAAAQAD	308
orf25.pep	RQEYAEYLKLCQDTRMTRERIQYLRGYSID	60
orf25ng	RQEYAEYLKLCQDTRMTRERIQYLRGYSID	338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGTATCGGA	AAC TCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCAGCGTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCCGCCGTGT
	101	TGCAGGACAT	ACGCGGCAGT	ATTCAGGAAA	CGCTCACGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCGG	ACAAAATTAT
	201	CGCCGCCGCC	TACGGTTTGG	CGTTTCTTT	GGAACACGCT	TCGGAACGCG
50	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTGTAACAT	TACCGTGCCG
	301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAAC	AGCCCCCTGC	TGTATGGGGA
	351	AACGTCTTTG	GCAGACATCG	TGCAGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAGACGCG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGCCAAAGAC
	451	GCTCGGACGG	CATTTATCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
	501	GTCTGCCGCG	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
55	551	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTTGAG	CGGCAAAGCC
	601	CGTGAAGAAG	AACCGTCCAA	ACCCACCCCC	GAAGACATTT	TGGAACACAA
	651	TGCGGCCGCG	GCGGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCACCCG
	701	AACCCGAAAT	CCTGCATCCC	GACGACGTCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AACGTGCGGA
60	801	ATCCGAAATT	ACCAAACCTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
	851	AGTTGGTCCG	CGAACAGCGC	AAGTGGGCGC	AGGAAAAAAT	CAGcaactgc
	901	cgACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCGG	AATACCTCAA
	951	GCTCCAATGC	GACACGCGGA	TGACGCGCGA	ACggaTACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

-393-

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR  
 51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP  
 101 SETLADAEAN SPLLYGETSL ADIVQKKTGG NVEFKDGVLT AAVRFLPAKD  
 151 ARTAFIDNTV GMATQTLASA LLPYGVKSIV MIDGKAVTKE DAVRVLSGKA  
 5 201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT  
 251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEOR KWAQEKISNC  
 301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID\*

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

10	orf25-1.pep	10	20	30	40	50	60
		MYRKLIALPFALLLAACGRE	EPPKALECANPAVLQ	GIRGNIQETLTQ	EARSFAREDGRQF		
	orf25ng	MYRKLIALPFALLLAACGRE	EPPKALECANPAVLQ	DIRGSIQETLTQ	EARSFAREDGRQF		
		10	20	30	40	50	60
15	orf25-1.pep	70	80	90	100	110	120
		VDADKIIAAAYGLAFSLEH	ASETQEGGRTFCIADLN	ITVPSETLADAKANSPL	LYGETAL		
	orf25ng	VDADKIIAAAYGLAFSLEH	ASETQEGGRTFCIADLN	ITVPSETLADAEANSPL	LYGETSL		
		70	80	90	100	110	120
20	orf25-1.pep	130	140	150	160	170	180
		SDIVRQKTGGNVEFKDGV	LTAAVRFLPVKDGQTAF	VNDTVGMAAQTLAALLP	YGVKSIV		
	orf25ng	ADIVQKKTGGNVEFKDGV	LTAAVRFLPAKDARTAF	IDNTVGMATQTLAALLP	YGVKSIV		
25		130	140	150	160	170	180
	orf25-1.pep	190	200	210	220	230	240
		MIDGKAVKKEDAVRILSG	KAREEPPSKPTPEDILE	HNAAGGDAGVPQAAEGA	PEPEILHP		
	orf25ng	MIDGKAVTKEDAVRVLSG	KAREEPPSKPTPEDILE	HNAAGGDAGVPQAAEGA	PEPEILHP		
30		190	200	210	220	230	240
	orf25-1.pep	250	260	270	280	290	300
		DDGERADTVTVSRGEVEE	EARVQNQRAESEITKLW	GGLDTPVQKELVGEOR	KWAQEKISNC		
	orf25ng	DDVERADTVTVSRGEVEE	EARVQNQRAESEITKLW	GGLDTPVQKELVGEOR	KWAQEKISNC		
35		250	260	270	280	290	300
	orf25-1.pep	310	320	330	339		
		RQAAAQADRQ EYAEYLK	LCQDTRMTRERIQYLR	GYSIDX			
	orf25ng	RQAAAQADRQ EYAEYLK	LCQDTRMTRERIQYLR	GYSIDX			
40		310	320	330			

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described  
 50 above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and  
 55 that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

### Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGwysGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
     201  CGsyGATTGG TCGCTGGGCA AACCAAAAAT CTTGGTTTTC CKGATACGTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....
      //
     851  .....AC TTCGCTGGTA
     901  TTCGGCGGCA CTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
     951  GATTA AAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
    1001  GTTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
    151  1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
     1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
     1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
     1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
     1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
    201  1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
     1351  GACCACGTTA CCTCGCAACT GCCTTAGGCC TTAACCGTTG CCGCCGCCGC
     1401  CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
     1451  TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
    1501  AAAAAA..

```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

30      1  MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFLVGNPNV
      51  DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSL LTYSGSN...
      //
     251  .....TSLV
     301  FGGTCGVFAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
     351  VGEMHTGDYL STL VAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
     401  IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNHI
     451  DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
     501  KK..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

40      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGTCCGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
     201  CGCGGATTGG TCGCTGGGCA AACCAAAAAT CTTGGTTTTC CTGATACGTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
     301  GCCGACTGGG CAAAACGGCA CATTA AAAAC CCGCGCGGCG CGAAAAATGCT
     351  GACCGCTGCG CTCGTGTTG TAACCTTTAT CGACGACTAT TTCCACAGTC
     401  TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
     451  CGCACCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT
     501  GCTGATGCCC GTTTC AAGCT GGGCGCGGTC GATTATCGCC ACGCTTGCCG
     551  GACTGCTCGT TACCTACAAA ATCACC GAAT ACACGCCGAT GGGGACGTTT
     601  GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
     651  GTTCGTCGTC GCATGGTTTT CCTTCGACAT CGGCTCGATG GCACGTTTCG
     701  AACAAAGCCG GTTGAACGAA GCCACGATG AAAGTCCGT TTCAGACGCT
     751  ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTTGG CCTTAATCGC
     801  CTCACACGTT TCCGCCATGA TCTACACGGG CGCGCAGGCA AGCGAAACCT
     851  TCAGCATTTT GGGGGCATTT GAAAACACGG ACGTAAACAC TTCGCTGGTA
     901  TTCGGCGGCA CTGCGGCGT CCTTGCCGTC GTTCTCTGCA CGCTCGGCAC
     951  GATTA AAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
    1001  GTTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
    1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
    1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
    1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
    1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
    1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
    1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

```



-395-

```

1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
1451 TTGGCAGCAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
1501 AAAAAACGCG CCAACGCCTG A

```

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

```

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
51 DGLTHLKDMV VGLAWSGDW SLGKPKILVF LILLGIFTSL LTYSGSNQAF
101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHSLAVGAIA RPVTKFKVS
151 RTKLAYILDS TAAPMCVLMV VSSWGASIIA TLAGLLVYK ITEYTPMGTF
10 201 VAMSLMNYA LFALIMVFVW AWFSDIGSM ARFEQAALNE AHDETAVSDA
251 TKGRVYALII PVLALIASTV SAMIYTGAAQ SETFSILGAF ENTVDNTSLV
301 EGGTCGVLA VCLTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
351 VGEMHTGDYL STLAVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI
15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTGIV LAVLIFLLKD
501 KKRANA*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H. influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

```

Orf26 1 MQLIDYSHSFFSVVPPFLALALAVITRRVXXXXXXXXXXVAFLVGGNPVDGLTHLKDMV 60
M+LID+S S +S+VP LA+ LA+ TRRV L +L V
HI1586 14 MELIDFSSSVWSIVPALLAIIAIAIATRRVLVSLSAGIIGSLMSLDWQIGSAFNVLVKNV 73

Orf26 61 VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSN 97
V L ++D + + I++F +LLG+ T+LLT SGSN
HI1586 74 VSLVYADGEIN-SNMNIVLFLLLGLVLTALLTVSGSN 109

//

Orf26 86 IFTSLTYSGS--NTSLVFGGTGCGVFAVVLCTL--GTIKTADYPKAVWQGAKSMTGXXXX 141
+F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G
HI1586 299 VFSVLGTGFENTVVGTSVLVVGFCSSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGSAIAI 358

Orf26 142 XXXXXXSTVVGEMHTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP 201
+ +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP
HI1586 359 LFFAWTINKIVGDMQTKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLP 418

Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISTTILSSTGARNHIDHVTSQXXXX 261
IAAAMA P L++PC+SAVMAGAVCGDHCS+SDTTILSSTGA+CNHIDHVT+Q
HI1586 419 IAAAMAANAPELLPCLSAVMAGAVCGDHCSFVSDTTILSSTGAKCNHIDHVTQLPYA 478

Orf26 262 XXXXXXXXXXXXXXXXSALLGFGTGIVLAVLIFLLKDK 302
S L GF T + L V+IF +K +
HI1586 479 ATVATATSIGIYIVVGFTYSGLAGFAATAVSLIVIFAVKKR 519

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N.*

*meningitidis*:

```

50 orf26.pep      10      20      30      40      50      60
      MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV
orf26a      MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
      10      20      30      40      50      60

orf26.pep      70      80      90      99
      VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSNXX-----
orf26a      VGLAWSGDWDLGKPKXLVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
      70      80      90      100      110      120

```

-396-

```

orf26.pep -----
5 orf26a  LVFVTFIDDYFHS LAVGAXARPVTDKFKVSRK LAYILDSTAAPMCVLMFVSSWGASIIA
      130      140      150      160      170      180

10 orf26.pep -----
orf26a  TLAGLLV TYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE
      190      200      210      220      230      240

15 orf26.pep -----
orf26a  AHDETA VSDGSGR VYAL IIPVLAL IASTV SAMIYTGAQASETFSILGAFENTDVNTSLV
      250      260      270      280      290      300
                                100      110
                                -----TSLV
                                ||||

20 orf26.pep 120      130      140      150      160      170
orf26a  FGGTCGVFAVVLCTLTG TIKTADY PKAVWQGA KSMFGA IAILILAWL ISTVVGEMHTGDYL
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf26a  FGGTCGVLA VVLCTLTG TIKIADY PKAVWQGA KSMFGA IAILILAWL ISTVVGEMHTGDYL
      310      320      330      340      350      360

25 orf26.pep 180      190      200      210      220      230
orf26a  STL VAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf26a  STL VAGNIHPGFLXVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVDP SLIIPCMSA
      370      380      390      400      410      420

30 orf26.pep 240      250      260      270      280      290
orf26a  VMAGAVCGDHCSPISD TILSSTGARC NHIDHVTSQLPYALT VAAAAASGYLALGLTKSA
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
35 orf26a  VMAGAVCGDHCSPISD TILSSTGARC NHIDHVTSQLPYALT VAAAAASGYLALGLTKSA
      430      440      450      460      470      480

orf26.pep 300      310
orf26a  LLGFGTTGIVLAVLIFLLKDKK
      |||||:|||||:|||||:|||||
40 orf26a  LLGFGXTGIVLAVLIFLLKDKKRANAX
      490      500

```

The complete length ORF26a nucleotide sequence <SEQ ID 693> is:

```

45 1 ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
51 51 TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
101 GCATCGGTAT TCTGGTCGGC GTTGCCCTTT TGGTCGGCGG CAACCCCGTC
151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTGGTCAAGA
201 CGGCGATTGG TCGCTGGGCA AACCAAAANT CTGGGTTTTC CTGATACTTT
251 TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
50 301 GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGGCG CGAAAATGCT
351 GACCGCCTGC CTCGTGTTCT TAACCTTTAT CGACGACTAT TTCCACAGTC
401 TCGCCGTCGG TCGNNTTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
451 CGCGCCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCGC CTATGTGCGT
55 501 GCTGATGCCC GTTCAAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
551 GACTGCTCGT TACCTACAAA ATCACC GAAT ACACGCCGAT GGGGACGTTT
601 GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
651 GTTCGTCTGC GCATGGTTCT CCTTCGACAT CGGCTCGATG GCACGTTTCG
701 AACCAAGCCG GTTGAACGAA GCCCACGATG AAAGTCCGCT TTCAGACGGC
751 AGCTGGGGCA GGGTTTACGC ATTGATTATT CCCGTTTGG CCTTAATCGC
60 801 CTCAACGGTT TCGCCATGA TCTACACCGG TGCACAGGCA AGCGAAACCT
851 TCAGCATTTT GGGTGCATTT GAAAATACGG ACGTGAACAC TTCGCTGTGA
901 TTCGGCGGCA CTGCGGCGT GCTTGCCGTC GTCCTCTGCA CGCTCGGCAC
951 GATTAAAATC GCCGATTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCCA
1001 TGTTGCGGCG AATCGCCATT TTAATCCTTG CCTGGCTCAT CAGTACGGTT
65 1051 GTCGGCGAAA TGCACACAGG CGACTACCTC TCCACGCTGG TTGCGGGCAA
1101 CATCCATCCC GGCTTCCTGN CCGTCATCCT TTCTCTGCTC GCCAGCGTGA
1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT CATGCTGCCG
1201 ATTGCCGCGC CCATGGCGGT CAAAGTCGAT CCCTCACTGA TTATCCCGTG
1251 TATGTCCGCC GTGATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
70 1301 TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

```

-397-

1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC  
 1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT  
 1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT  
 1501 AAAAAACGCG CCAACGCCTG A

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV  
 51 DGLTHLKDMV VGLAWSGDW SLGKPKXLVF LILLGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDDY FHSLAVGAXA RPYTDKFKVS  
 151 RAKLAYILDS TAAPMCVLMV VSSWGASIIA TLAGLLVITYK ITEYTPMGTF  
 10 201 VAMSLMNYA LFALIMVFVW WFSFDIGSM ARFEQAALNE AHDEAVSDG  
 251 SWGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV  
 301 FGGTCGVLAV VLCTLTGTIKI ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
 351 VGEHMTGDYL STLAVAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP  
 401 IAAAMAVKVD PSLIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI  
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD  
 501 KKRANA\*

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

		10	20	30	40	50	60
20	orf26a.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGV	VAFLVGGNPVDGLTHLKDMV				
	orf26-1	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGV	VAFLVGGNPVDGLTHLKDMV				
		10	20	30	40	50	60
		70	80	90	100	110	120
25	orf26a.pep	VGLAWSGDGWSL	GKPKXLVFLILLGIFTSL	LTYSGSNQAFADWAKRHIKNRRGAKMLTAC			
	orf26-1	VGLAWSGDGWSL	GKPKILVFLILLGIFTSL	LTYSGSNQAFADWAKRHIKNRRGAKMLTAC			
		70	80	90	100	110	120
30	orf26a.pep	LVFVTFIDDYFHSLAVGAXARPYTDKFKVSR	AKLAYILDSTAAPMCVLMV	VSSWGASIIA			
	orf26-1	LVFVTFIDDYFHSLAVGAIARPYTDKFKVSRT	KLAYILDSTAAPMCVLMV	VSSWGASIIA			
		130	140	150	160	170	180
35	orf26a.pep	TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVW	WFSFDIGSMARFEQAALNE				
	orf26-1	TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVW	WFSFDIGSMARFEQAALNE				
		190	200	210	220	230	240
40	orf26a.pep	AHDEAVSDGSWGRVYALII	PVLALIASTVSAMIYTGAQASE	TFSILGAFENTDVNTSLV			
	orf26-1	AHDEAVSDATKGRVYALII	PVLALIASTVSAMIYTGAQASE	TFSILGAFENTDVNTSLV			
		250	260	270	280	290	300
45	orf26a.pep	FGGTCGVLAVVLCTLTGTIKIADYPKAVWQGA	KSMFGAIAILILAWLISTV	VGEHMTGDYL			
	orf26-1	FGGTCGVLAVVLCTLTGTIKTADYPKAVWQGA	KSMFGAIAILILAWLISTV	VGEHMTGDYL			
		310	320	330	340	350	360
50	orf26a.pep	STLVAGNIHPGFLXVILFLLASVMAFATGTSWGTFGIMLP	IAAAMAVKVDPSLIIPCMSA				
	orf26-1	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP	IAAAMAVKVEPALIIPCMSA				
		370	380	390	400	410	420
55	orf26a.pep	VMAGAVCGDHCSPI	SDTTILSSTGARNHIDHVTSQ	LPYALTVA	AAAAASGYLALGLTKSA		
	orf26-1	VMAGAVCGDHCSPI	SDTTILSSTGARNHIDHVTSQ	LPYALTVA	AAAAASGYLALGLTKSA		
		430	440	450	460	470	480
60	orf26a.pep	LLGFGXTGIVLAVLIFLLKDKKRANAX					
	orf26-1	LLGFGXTGIVLAVLIFLLKDKKRANAX					
		490	500				
65	orf26a.pep	LLGFGXTGIVLAVLIFLLKDKKRANAX					

-398-

```

      |||||:|||||
orf26-1  LLGFGTTGIVLAVLIFLLKDKKRANAX
              490       500

```

### 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

```

10 orf26.pep  MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFVLVGGNPVDGLTHLKDMV 60
    orf26ng  MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVLVGGNPVDGLTHLKDMV 60

    orf26.pep  VGLAWSDXDWSLGPVKILVFXILLGIFTSLLTYSGSN 97
    orf26ng  VGLAWADGDWSLGPVKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKRCGAKMLTAC 120

15 //

    orf26.pep  TSLVFGGTCGVFAVVLCTLGTIKTADYPKA 326
    orf26ng  ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTCGVLAIVLCTFGTIKTADYPKA 326

20 orf26.pep  VWQGAKSMFGAIAILILAWLISTVVGEMHTGDYLSLTVAGNIHPGFLPVILFLLASVMAF 386
    orf26ng  VWQGAKSMFGAIAILILAWLISTVVGEMHTGDYLSLTVAGNIHPGFLPVILFLLASVMAF 386

25 orf26.pep  ATGTSWGTFGIMLPAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR 446
    orf26ng  ATGTSWGTFGIMLPAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR 446

30 orf26.pep  CNHIDHVTSQLPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK 502
    orf26ng  CNHIDHVTSQLPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKKRADV 506

```

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

```

35 1 ATGCAGCTGA TTGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
    51 TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
    101 GCATCGGTAT TTTGGTCGGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
    151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGGCAGA
    201 CCGCGATTGG TCGCTGGGCA AACCAAAAAT CTGGGTTTTT CTGATACTTT
    251 TGGGCATTTT CACTTCACTG CTGACCTACT CCGGCAGCAA TCAGGCGGTT
    301 GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGTGCGGCG CGAAAATGCT
    351 GACGCCTGCG CTCGTGTTTC TAACCTTTAT CGACGACTAT TTCCACAGCC
    401 TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
    451 CGCGCCAAAC TCGCCTACAT CCTCGACTCC ACTGCCTCGC CCATGTGCGT
    501 GCTGATGCCC GTTTCAGCTT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
    551 GATTGCTCGT TACCTACAAA ATTACCGAAT ACACGCCGAT GGGGACGTTT
    601 GTCGCCATGA GCCTGATGAA CTATTACGCG CTGTTTGCCC TGATTATGGT
    651 ATTCGTCGTC GCATGGTTCT CCTCGACAT CGGCTCGATg gCGCGTTTCG
    701 AACAGGCTGC GTTGAACGAA gcccaggacg aaaccgcccgc tTCAGACgCT
    751 ACCAAAGGTC GTGTTTACGC ATTGATTATT CCCGTTTTGG CCTTAATCGC
    801 CTCAACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
    851 TCAGCATTTT GGGGCGATT GAAATACCG ACGTAAACAC TTCGCTGGTA
    901 TTCGGCGGCA CTTGCGGCGT GCTTGCCGTC GTCTCTGCA CGTTCGGCAC
    951 GATTAAAACC GCCGATTATC CCAAAGCCGT GTGGCAGGGT GCGAAATCCA
    1001 TGTTGCGGCG AATCGCCATT TTAATCCTCG CCTGGCTCAT CAGTACGGTT
    1051 GTCGGCGAAA TGACACAGGG CCACTACCTC TCCACGCTGG TTGCGGGCAA
    1101 CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
    1151 TGGCGTTTTC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
    1201 ATTGCGCGCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTAtcccGTG
    1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CCGCGACACC TGTTGCCCCA
    1301 TCTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACACATC
    1351 GACCACGTTA CCTCGCAACT GCCTTATGCC CTGACGGTTG CCGCGCCGCG
    1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
    1451 TTGGCACGAC CGGTATTGTA TTGGCGGTGC TGATTTTCTT GTTGAAAGAT
    1501 AAAAACGCGC CCGACGTTTG A

```

65 This encodes a protein having amino acid sequence <SEQ ID 696>:

-399-

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGGNPV  
 51 DGLTHLKDMV VGLAWADGDW SLGKPKILVF LILLGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RCGAKMLTAC LVFVTFIDDY FHSLAVGAIARPVTDFKFKVS  
 5 RAKLAYILDS TASPMLVMP VSSWGASIIA TLAGLLVITYK ITEYTPMGTF  
 201 VAMSLMNYIA LFALIMVFVW AWFSFDIGSM ARFEQAALNE AQDETAASDA  
 251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV  
 301 FGGTCGVLAV VLCTFGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
 351 VGEMHTGDYL STLAVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP  
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH  
 10 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD  
 501 KKRADV\*

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

15	orf26-1.pep	10	20	30	40	50	60
		MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLVGGNPV	DGLTHLKDMV
	orf26ng	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLVGGNPV	DGLTHLKDMV
20	orf26-1.pep	70	80	90	100	110	120
		VGLAWSGDW	SLGKPKILVF	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RRCGAKMLTAC
	orf26ng	VGLAWADGDW	SLGKPKILVF	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RRCGAKMLTAC
25	orf26-1.pep	130	140	150	160	170	180
		LVFVTFIDDY	FHSLAVGAIAR	PVTDFKFKVS	RKTALAYILD	STAAPMCVLM	PVSSWGASIIA
	orf26ng	LVFVTFIDDY	FHSLAVGAIAR	PVTDFKFKVS	RKTALAYILD	STAAPMCVLM	PVSSWGASIIA
30	orf26-1.pep	190	200	210	220	230	240
		TLAGLLVITYK	ITEYTPMGTF	VAMSLMNYIA	LFALIMVFVW	AWFSFDIGSM	ARFEQAALNE
	orf26ng	TLAGLLVITYK	ITEYTPMGTF	VAMSLMNYIA	LFALIMVFVW	AWFSFDIGSM	ARFEQAALNE
35	orf26-1.pep	250	260	270	280	290	300
		AHDETAVS	DATKGRVYALII	PVLALIASTV	SAMIYTGAQA	SETFSILGAF	ENTDVNTSLV
	orf26ng	AQDETAAS	DATKGRVYALII	PVLALIASTV	SAMIYTGAQA	SETFSILGAF	ENTDVNTSLV
40	orf26-1.pep	310	320	330	340	350	360
		FGGTCGVLAV	VLCTFGTIKT	ADYPKAVWQG	AKSMFGAIAI	LILAWLISTV	VVGEMHTGDYL
	orf26ng	FGGTCGVLAV	VLCTFGTIKT	ADYPKAVWQG	AKSMFGAIAI	LILAWLISTV	VVGEMHTGDYL
45	orf26-1.pep	370	380	390	400	410	420
		STLVAGNIHP	GFLPVILFLL	ASVMAFATGT	SWGTFGIMLP	IAAAMAVKVE	PALIIIPCMSA
	orf26ng	STLVAGNIHP	GFLPVILFLL	ASVMAFATGT	SWGTFGIMLP	IAAAMAVKVE	PALIIIPCMSA
50	orf26-1.pep	430	440	450	460	470	480
		VMAGAVCGDH	CSPISDTTIL	SSTGARCNIH	DHVTSQLPYA	LTVAAAAASG	YLALGLTKSA
	orf26ng	VMAGAVCGDH	CSPISDTTIL	SSTGARCNIH	DHVTSQLPYA	LTVAAAAASG	YLALGLTKSA
55	orf26-1.pep	490	500				
		LLGFGTTGIV	LAVLIFLLK	DKKRANAX			
	orf26ng	LLGFGTTGIV	LAVLIFLLK	DKKRADVX			
60	orf26-1.pep	490	500				
		LLGFGTTGIV	LAVLIFLLK	DKKRADVX			
65	orf26ng	490	500				

In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

-400-

sp|P44263|YF86\_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir||C64037  
 hypothetical  
 protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.  
 influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519  
 Score = 538 bits (1370), Expect = e-152  
 Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXAFLVGGNPVDGLTHLKMV 60  
 M+LID+S S +S+VP LA+ LA+ TRR L +L V  
 Sbjct: 14 MELIDFSSSVWSIVPALLAII LAIATRRVLVLSAGIIGSIMLSDWQIGSAFNVLKVN 73

Query: 61 VGLAWADGDWSLKGPKILVFLILLGIFTSLITYSGSNQAFADWAKRHIKNRCGAKMLTAC 120  
 V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A  
 Sbjct: 74 VSLVYADGEIN-SNMNIVFLLLLLGVLTALLTVSGSNRAFAEWAQSRIKGRGAKLLAAS 132

Query: 121 LVFVTFIDDYFHS LAVGAIARPVTDKFKVSRK LAYILDSTASPMCVIMPVSSWGASIIA 180  
 LVFVTFIDDYFHS LAVGAIARPVTD+FKVSRK LAYILDSTA+PMCV+MPVSSWGA II  
 Sbjct: 133 LVFVTFIDDYFHS LAVGAIARPVTDREKFSRAK LAYILDSTAAPMCVMMPVSSWGAYIIT 192

Query: 181 TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFFVAVWFSFDIGSMARFEQAALNE 240  
 + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL  
 Sbjct: 193 LIGGLLATYSITEYTPIGAFVAMSSMNFYAIFSIIMVFFVAYFSFDIASMVRHEKLALKN 252

Query: 241 AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAAQ----SETFSILGAFENTDVN 296  
 +D+ TKG+V LI+P+L LI +TVS MIYTGAA+ + FS+LG FENT V  
 Sbjct: 253 TEDQLEETGTGKQVRNLILPILVLIATVSMMIYTGAELAADGKVFVSVLGTFFENTVVG 312

Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWQGA KSMFGXXXXXXXXXXSTVVGEM 354  
 TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M  
 Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGAIILFFFAWTINKIVGDM 372

Query: 355 HTGDYLS TLVAGNIHPGFLPVILFLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALI 414  
 TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP IAAAMA P L+  
 Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLP IAAAMAANAPELL 432

Query: 415 IPCMSAVMAGAVCGDHCSPISD TTILSSTGARC NHIDHVT SQXXXXXXXXXXXXXXXXX 474  
 +PC+SAVMAGAVCGDHCSP+SD TTILSSTGA+CNHIDHVT+Q  
 Sbjct: 433 LPCLSAVMAGAVCGDHCSPVSD TTILSSTGAKCNHIDHVTQLPYAATVATATSIGYIVV 492

Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501  
 S L GF T + L V+IF +K +  
 Sbjct: 493 GFTYSGLAGFAATAVSLIVIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,  
 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA  
 51 CGATGAGCCT GCCAAATTC TGA CTGGGA TGAAAGCGGC CGATTACTCT  
 101 CGGA ACTGTC TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG  
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT  
 201 GGTCAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCCCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW  
 51 YEDGSKKSEX VYQDDKLVRK TQWDKDG YLI EP\*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

1 ATGAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC  
 51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA  
 101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG  
 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA

-401-

201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC  
 251 CTACCCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA  
 301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA  
 351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGCT  
 5 401 TGAGTGAGGG TACGGGATAC CGCTATTACC GTAACGGCGG CAAGGAAAGC  
 451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAAC GGCATATGGA AGCAATGGTA  
 501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG  
 551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT  
 10 601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG  
 651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA  
 701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCTGA

This corresponds to the amino acid sequence <SEQ ID 700; ORF27-1>:

1 MKKLSRIVFS TVLLGFSAAL PAQYTSVYFN QNGKLTATMS SAAYIROYSV  
 51 VAGIAHAQDF YPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK  
 15 101 KMAGGFSK GK PDGEWVNWYP NGKKSVMMPY KNGLSEGTGY RYYRNGGKES  
 151 EIQFKQNKAN GVWKQYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS  
 201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

					10	20	30
	orf27.pep				KQWYADXS	SIKTEMVMVN	DEPAKILTWD
						:	
25	orf27a	LSEGTGXRYR	NGGKESEI	QFKQNKANG	VWKQWYADG	NIKTEMVMVN	DEPAKILTWD
		140	150	160	170	180	190
		40	50	60	70	80	
	orf27.pep	RLLSELSIR	HHQRNGVV	LEWYEDGSKK	SEXVYQDDK	LVRKTQWDK	DGYLIEPX
			:				
30	orf27a	RLLSELSI	HHXRNGVV	LEWYEDGSKK	XEAVYQDDK	LVRKTQWDK	XDGYLIEPX
		200	210	220	230	240	

The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCTGT TGGGTTTTTC  
 35 51 GGCCGCTTTG CCGGCGCAGA NCTATTCTGT TTATTTTAAT CAGAACGGGA  
 101 AACTGACGGC GACGNTGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG  
 151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA  
 201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC  
 40 251 CTACCCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGGTCAGAAA  
 301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA  
 351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGTT  
 401 TGAGTGAAGG TACGGGGTNN CGCTATTACC GTAACGGCGG CAAGGAAAGC  
 451 GAAATCCAGT TTAACAGAA TAAGGCAAAAC GGCATATGGA AGCAATGGTA  
 501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG  
 45 551 CCAAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAAGTGTCT  
 601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG  
 651 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA  
 701 CCCAGTGGGA TAANGATGGT TATTTAATCG AACCTGA

This encodes a protein having amino acid sequence <SEQ ID 702>:

1 MKKLSRIVFS TVLLGFSAAL PAQXYSVYFN QNGKLTATXS SAAYIROYSV  
 51 AEGIAHAQXF XYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK  
 101 KMAGGFSK GK PDGEWVNWYP NGKKSVMMPY KNGLSEGTGX RYYRNGGKES  
 151 EIQFKQNKAN GVWKQYADG NIKTEMVMVN DEPAKILTWD ESGRLLSELS  
 201 IHHHRNGVV LEWYEDGSKK XEAVYQDDKL VRKTQWDKDG YLIEP\*

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

		10	20	30	40	50	60
	orf27a.pep	MKKLSRIVF	STVLLGFSAAL	PAQXYSVYFN	QNGKLTATXS	SAAYIROYSV	AEGIAHAQXF

[illegible]

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from

```

      orf27.pep                                KQWYADXSIKTEMVMVNDEPAKILTWDESG    30
                                         ||||| ||||| ||||| ||||| |||||
      orf27ng          LSEGTYGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDESG    193
35      orf27.pep          RLLSEL SIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDGYLIEP        82
                                         ||||| :|||: ||||| ||||| ||||| |||||
      orf27ng          RLLSEL SIRHHRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDGYLIEP        245

```

40	1	ATGAAGAAAT	TATCTCGGAT	TGTATTTTCA	ATCGTACTGT	TGGGTTTTTC
	51	GGCCCGCTTTG	CCGGCGCAGA	CCTATTCTGT	TTATTTTAAAT	CAGAACGGGA
	101	AACTGACGGC	GACGATGTCT	TCTGCCGCTT	ATATCAGGCA	ATATAGTGTG
	151	CGGGCGGGTA	TGCACACGC	GCAGGATTTT	TATTATCCGT	CGATGAAGAA
45	201	ATATTCGGAA	CCTTATATCG	TFGCTTCAAC	GCAAATCAA	TCTTTTGTGC
	251	CTACCCGTGCA	AAACGGTATG	TTGATTTTGT	GGCATTTTAA	TGGTCAAGAA
	301	AAAATGGCG	GGGGCTTACG	CAAGGGTAA	CCGGACGGG	AAATGGGTCAA
	351	CTGGTATCCG	AACCGTAAAA	AATCTGCGGT	TATGCCTTAT	AAAAATGGCT
50	401	TGAGTGAGGG	TACGGGATAC	CGTTATTACC	GTAACGGCGG	CAAGGAAAGC
	451	GAAATCCAGT	TTAAGCAAAA	TAAGGCGCAAC	GGCGTATGGA	AGCAATGGTA
	501	TGCCGATGGA	AGTATCAAGA	CGGAAATGGT	TATGGTCAAC	GATGAGCCGT
	551	CCAAAATTCT	GACTTGGGAT	GAAAGCGGCC	GATTACTTTC	GGAACTGTCT
	601	ATCCGCACC	ATAAACGCAA	CGGGGTGGTT	TTGGATTTGGT	ATGAAGATGG
	651	TTCTAAAAAG	AGCGAGGCTG	TTTATCAGGA	TGACAAGTTG	GTCAGGAAAA
	701	CCCAATGGGA	TAAGGATGGT	TATTTAATCG	AACCCTGA	

55

1	MKKLSRIVFS	IVLLGFSAAAL	PAQTYSVYFN	QNGKLTATMS	SAAYIRQYSV
51	AAGIAHAQDF	YYPSPMKKYE	PYIVASTQIK	SFVPTLQNGM	LILWHFNGQK
101	KMAGGFSK GK	PDGEWNWYP	NGKKSAMVPI	KNLSEGTGY	RYRRNGKES
151	EIQFKQNKAN	GWVKQWYADG	SIKTEVMVMN	DEPAKLTITWD	ESGRLLSELS
201	IRHHKRGV	LVWYEDGSKK	SEAVYODDKL	VRKTOWDKDG	YLIEP*

```

              10      20      30      40      50      60
orf27-1.pep  MKKLSRIVFSTVLLGFSAALPAQTYSVYFNONGKLTATMSSAAYIRQYSVVAGIAHAADF

```



								:		
	orf27ng	MKKLSRIVFSI	VLLGFSAALPAQ	TYSVYFNGKLTATMSSAAYIRQYSVAAGIAHAQDF						
		10	20	30	40	50	60			
5		70	80	90	100	110	120			
	orf27-1.pep	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNQGKKMAGGFSKGKPDGEWVNWYP								
	orf27ng	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNQGKKMAGGFSKGKPDGEWVNWYP								
10		70	80	90	100	110	120			
		130	140	150	160	170	180			
	orf27-1.pep	NGKKSAVMPYKNGLSEGTYRYRRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN								
15	orf27ng	NGKKSAVMPYKNGLSEGTYRYRRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN								
		130	140	150	160	170	180			
		190	200	210	220	230	240			
	orf27-1.pep	DEPAKILTWEESGRLLSEL SIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKD G								
20										
	orf27ng	DEPAKILTWEESGRLLSEL SIRHHRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKD G								
		190	200	210	220	230	240			
25	orf27-1.pep	YLIEPX 								
	orf27ng	YLIEPX								

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

50 This corresponds to the amino acid sequence <SEO ID 706; ORF47>:

1 MKFTKHPVWA MAFRFYSLA ALYGALSULL WGFYGTGTHX LSGFYWHAHE  
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTIFWL AARIAAFIPG  
101 WGASASGILG TLEFWYGAVC MALPVIRSON ORNYVAVEAL FVLGGTHAAF

-404-

151 HVQLHNGNLG GLLSGLQSGL VM

Further work revealed the complete nucleotide sequence &lt;SEQ ID 707&gt;:

```

      1  ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
      51  TTCGCTGGCG GCTCTGTACG GCGCATGTGC CGTATTGCTG TGGGGTTTCG
5      101  GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
      151  ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
      201  CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGGCGGC GTTCTGGTCG
      251  GCTTGACTAT CTTTGGCTG GCTGCGCGGA TTGCCGCTT TATCCCGGGT
      301  TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG
10     351  CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGAGAAT CAACGCAACT
      401  ATGTTGCCGT GTTCGCGCTG TTCGTCTGG GCGGCACGCA TCGGCGGTTT
      451  CACGTCCAGC TGCACAACGG CAACCTAGGC GGA CTCTGA GCGGATTGCA
      501  GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
      551  TTATTTCGTT TTTTACGTCC AAACGCTTGA ATGTGCGCA GATTCCAGT
15     601  CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCATGCG TGACTGCCAT
      651  GCTGATGGCG CACGGTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCTTTG
      701  CGGCAGGTGT GATTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAAGCC
      751  GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
      801  CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCCGCTTTCC
20     851  TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
      901  TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CATACGGGCA ATCCGATTTA
      951  TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
100    1001  CCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
105    1051  AGCATCCGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
25     1101  GTGGAAGTAT ATTCCTTGGC TGATTCTGCC GCGTTCGGAC GGCAGGCCCC
1151   GTTGA

```

This corresponds to the amino acid sequence &lt;SEQ ID 708; ORF47-1&gt;:

```

      1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
      51  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG
30     101  WGASASGILG TLEFFWYGAVC MALPVIRSQN QRNYVAVFAL FVLGGTHAAF
      151  HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GTRIIISFFTS KRLNVPQIPS
      201  PKWVAQASLW LPMLTAMLMA HGVLAWLSAV FAFAAGVIFT VQVYRWYKIP
      251  VLKEPMLWIL FAGYLFITGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
35     301  LGMMARTALG HTGNPIYPPP KAVPVAEWLM MAATAVRMVA VFSSGTAYTH
      351  SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 *meningitidis*:

```

      10      20      30      40      50      60
orf47.pep  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHXLSGFYWHAHEMIWGYAGLVV
      |||
orf47a     MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHLSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60

      70      80      90     100     110     120
orf47.pep  IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      |||
orf47a     IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      70      80      90     100     110     120

      130     140     150     160     170
orf47.pep  MALPVIRSQNQRNYVAVFALFVLGGTHAAAFHVQLHNGNLGGLLSGLQSGGLVM
      |||
orf47a     MALPVIRSQNQRNYVAVFALFVLGGTHAAAFHVQLHNGNLGGLLSGLQSGGLVMVSGFIGLI
      130     140     150     160     170     180

orf47a     GTRIIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFT
      190     200     210     220     230     240

```

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```
1  ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTTA
5  51  TTCACTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
    101  GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
    151  ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
    201  CGTCGCCACT TGGACGGGGC AGCCGCCCAC GCGGGGCGGC GTTCTGGTCG
    251  GCTTGACTAT CTTTGGGCTG GCTGCGCGGA TTGCCGCCCT TATCCCGGGT
    301  TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
    351  CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAATT
    401  ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGTACGCA CGCGGCGTTC
    451  CACGTCCAGC TGCACAACGG CAACCTAGGC GGA CTCTGA GCGGATTGCA
    501  GTCGGGCTTG GTGATGGTGT CGGGTTTAT CGGTCTGATT GGTACGCGGA
    551  TTATTTTCGT TTTTACGTCC AAACGGTTGA ATGTGCCGCA GATTCCCAGT
    601  CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGACCGCCAT
    651  GCTGATGGCG CACGGCGTGA TGCTTGGCT GTCGGCGGCT TTCGCGTTTG
    701  CGGCAGGTGT GATTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT
    751  GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
    801  CGGATTGGGG CTGATTGCGG TCGCGCGGTC TTATTTCAA CCGCTTTTCC
    851  TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
    901  TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATCCGATTTA
    951  TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
    1001  CCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
    1051  AGCATACGCA CCTCTTCGGT TTTGTTTGCA CTGCGGCTT TGGTGTATGC
    1101  GTGGAAGTAT ATTCCTTGGC TGATTCGTCC GCGTTCGAC GGCAGGCCCG
    1151  GTTGA
```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```
1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
5  51  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG
    101  WGASASGILG TLFYWGAVC MALPVIRSQN QRNYVAVFAL FVLGGTHAAF
    151  HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GTRIISFFTS KRLNVPQIPS
    201  PKWVAQASLW LPMLTAMLMA HGVMPWLSAA FAFAGVIFT VQVYRWYKYP
    251  VLKEPMLWIL FAGYLEFTGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT
    301  LGMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
    351  SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*
```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

```
10      20      30      40      50      60
orf47a.pep  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
40  orf47-1    MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60
      70      80      90     100     110     120
orf47a.pep  IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
45  orf47-1    IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      70      80      90     100     110     120
      130     140     150     160     170     180
orf47a.pep  MALPVIRSQNQRNYVAVFALEFVLGGTHAAHFVQLHNGNLGGLLSGLQSLGVMVSGFIGLI
50  orf47-1    MALPVIRSQNQRNYVAVFALEFVLGGTHAAHFVQLHNGNLGGLLSGLQSLGVMVSGFIGLI
      130     140     150     160     170     180
      190     200     210     220     230     240
orf47a.pep  GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFT
55  orf47-1    GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVLAWSAVFAFAAGVIFT
      190     200     210     220     230     240
      250     260     270     280     290     300
orf47a.pep  VQVYRWYKYPVLKEPMLWILFAGYLEFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
60  orf47-1    VQVYRWYKYPVLKEPMLWILFAGYLEFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
      250     260     270     280     290     300
      310     320     330     340     350     360
```

-406-

5	orf47a.pep	LGM MARTALGHTGNPIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
	orf47-1	LGM MARTALGHTGNPIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
		310 320 330 340 350 360
10	orf47a.pep	LALLVYAWKYIPWLIRPRSDGRPGX
	orf47-1	LALLVYAWKYIPWLIRPRSDGRPGX
		370 380

Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHAMIWGYAGLVV	60
	ORF47ng	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHAMIWGYAGLVV	60
20	ORF47	IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC	120
	ORF47ng	IAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC	120
25	ORF47	MALPVIRSONRNYVAVFAFVLGGTHAAAFHVQLHNGNLGGLLSGLQSGGLVM	172
	ORF47ng	MALPVIRSONRRNYVAVFAFVLGGTHAAAFHVQLHNGNLGGLLSGLQSGGLVMVWGFGLI	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

30	1	<u>MKFTKHPVWA</u>	<u>MAFRPFYSLA</u>	<u>ALYGALSVLL</u>	<u>WGFYGTGTHE</u>	<u>LSGFYWHAE</u>
	51	<u>MIWGYAGLVV</u>	<u>IAFLLTAVAT</u>	<u>WTGQPPTRGV</u>	<u>VLVGLTAFWL</u>	<u>AARIAAFIPG</u>
	101	<u>WGAAASGILG</u>	<u>TLFFWYGAVC</u>	<u>MALPVIRSON</u>	<u>RRNYVAVFAI</u>	<u>FVLGGTHAAAF</u>
	151	<u>HVQLHNGNLG</u>	<u>GLLSGLQSGL</u>	<u>VMVWGFGLI</u>	<u>GKIIISFFTS</u>	<u>KRLKLPQIPS</u>
	201	<u>PKWVAHASLW</u>	<u>LPMLNAILMA</u>	<u>HRVMPWLSAA</u>	<u>FPFAAGVIFT</u>	<u>VQVYAGGITP</u>
	251	<u>IEETSCGSVA</u>	<u>GICYRLGNSS</u>	<u>G</u>		

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

40	TM segments in ORF47ng			
	INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
	INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
	INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
	INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
	INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
	INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1	ATGAAATTTA	CCAAACATCC	CGTCTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCAC TGCGG	GCACTGTACG	GCGCATGTGC	CGTATTGCTG	TGGGGTTTCG
50	101	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTGGG	GTTATGCCGG	TCTCGTCGTC	ATCGCCTTCC	TGCTGACCGC
55	201	CGTCGCCACT	TGGACGGGAC	AGCCGCCAC	GAGGGCGCGG	GTTCTGGTCG
	251	GCTTGACCGC	CTTTTGGCTG	GCTGCGCGGA	TTGCCGCCTT	TATCCCGGGT
	301	TGGGGTGCGG	CGGCAAGCGG	CATACTCGGT	ACGCTGTTTT	TCTGGTACGG
	351	CGCGGTGTGC	ATGGCTTTGC	CCGTTATCCG	TtcgCAAAAC	CGGCGCAACT
	401	ATGtcgCCGT	ATTTCGAATA	TTTGTGCTGG	GCGGTACGCA	TGCGGcgTTC
	451	CACGtcAgc	tGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
	501	GTCGGGCGCTG	GTTATGGTGT	CGGGCTTTAT	CGGCCTGATT	GGGATGAGGA
	551	TTATTTTCGTT	TTTTACGTCC	AAACGGTTGA	ACGTGCCGCA	GATTCCCACT
	601	CCGAAATGGG	TGGCGCAGGC	TTCTGCTGTG	CTACCCATGC	TGACCGCCAT

-407-

5  
10  
1151

```

651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
701 CGGCGGGCGT GATTTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCTGCCTTCC
851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATTCGATTTA
951 TCCGCCGCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGGCAA
1001 CCGCCGTCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC
1101 GTGGAATAC ATTCCGTGGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

15  
20

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
51  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTAFWL AARIAAFIPG
101 WGAASGILG TLFFWYGAVC MALPVIRSON RRNYVAVFAI FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GMRIISFFTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAILMA HGVMPWLSAA FAFAGVIFT VQVYRWYKYP
251 VLKEPMLWIL FAGYLETGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
301 LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60

```

              10      20      30      40      50      60
orf47-1.pep  MKFTKHPVWMAFAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV
orf47ng-1    MKFTKHPVWMAFAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV
              10      20      30      40      50      60

              70      80      90     100     110     120
orf47-1.pep  IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
orf47ng-1    IAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC
              70      80      90     100     110     120

              130     140     150     160     170     180
orf47-1.pep  MALPVIRSONRRNYVAVFAIFVLGGTHAAAFHVQLHNGNLGGLLSGLQSLGVMVSGFIGLI
orf47ng-1    MALPVIRSONRRNYVAVFAIFVLGGTHAAAFHVQLHNGNLGGLLSGLQSLGVMVSGFIGLI
              130     140     150     160     170     180

              190     200     210     220     230     240
orf47-1.pep  GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAGVLAWSAVFAFAAGVIFT
orf47ng-1    GMRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAILMAHGVMPWLSAAFAFAAGVIFT
              190     200     210     220     230     240

              250     260     270     280     290     300
orf47-1.pep  VQVYRWYKYPVLKEPMLWILFAGYLETGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT
orf47ng-1    VQVYRWYKYPVLKEPMLWILFAGYLETGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT
              250     260     270     280     290     300

              310     320     330     340     350     360
orf47-1.pep  LGMMARTALGHTGNPIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
orf47ng-1    LGMMARTALGHTGNPIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
              310     320     330     340     350     360

              370     380
orf47-1.pep  LALLVYAWKYIPWLIRPRSDGRPGX
orf47ng-1    LALLVYAWKYIPWLIRPRSDGRPGX
              370     380

```

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

65 gnl|PID|e246540 (Z73914) ORF396 protein [*Pseudomonas stutzeri*] Length = 396  
Score = 155 bits (389), Expect = 5e-37

-408-

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

Query: 7 PVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59  
 P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +  
 5 Sbjct: 14 PIWRLAFRPFFLAGSLYALLAIPLVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71

Query: 60 VIAFLLTAVATWTGQPPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119  
 V FLLTAV TWTGQ G LVGL A WLAAR+ ++ G AA L LF  
 10 Sbjct: 72 VAGFLLTAVQTWTGTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPLDLLFLVALVW 130

Query: 120 CMALPVIRSONRRNYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVSQFVGL 179  
 MA + + +RNY V + ++ G +V+ + L  
 15 Sbjct: 131 MMAQMLWAVRQKRNYPIVVVLSMLGADVLIITGLLQNDALQRQGVLAGLWLVAALMAL 190

Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV---MPWLSAAFAFA 234  
 IG R+I FFT + L P W+ A L + A+L A GV P L F A  
 20 Sbjct: 191 IGGRVIPFFTQRLGKVDVAVKPVWLDVALLVGTGVIALHAFGVAMRPQPLGLLEFV-A 249

Query: 235 AGVIFTVQVYRWYKPVLPKEPMLWILFAGYLTGLGLIAGVASYF-KPAFXXXXXXXXXXX 293  
 GV +++ RW+ K + K +LW L L+ + + +F A  
 25 Sbjct: 250 IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLWLVAAFGLALWHFGLLAQSSPSLHALSV 309

Query: 294 XXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXFSSGTAYTHSIR 353  
 M+AR LGHTG + P + AF L F S +  
 30 Sbjct: 310 GSMSGLILAMIARVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAWPVGGWL 365

Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384  
 ++V + LA +Y W+Y P L+ R DG PG  
 35 Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

35 1 ..ATGCCGCTCTG AAGGTTTCAGA CGGCmTCGGT GyCGGGGAay CAGAAGyGGT  
 51 AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG  
 101 CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT  
 151 GATGTCGAaA CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC  
 201 TGtKGCTTTC GTGATAGGsA GGTtGyTGG kmksAsyTTG TAyrATwkkG  
 40 251 CCTssCwsTG kAGmGCCkTk CkyTGGTkkA swGrwArTAG TCGTGGTtTy  
 301 TkTtYyCACC GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG  
 351 CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG  
 401 GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT  
 45 451 TGTTTGGGTT TCTTTGTAGT TGTtGTTTAT CTCTTCAGTA ACTTTTTTAG  
 501 TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT  
 551 ATTCTCCAGC CGCCGAAATC ..

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

50 1 ..MPSESGDXG XGEXEXVAHA QXDFVGFEAG VFQASPVVVT VSGVXXQLGX  
 51 DVETDTGDDT KTXAADXVAF VIGRFxGXXL YXXAXXXAX XWXXXXSRGF  
 101 XXHRMNLMFN VSVGDARADI GFEFIVEFEI VNGGQAERRN GVEAAVSLMF  
 151 CLGFFVVVVY LFSNFFSRI TFFPFSVTGI ICRYSPAEEI ..

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

-409-

	orf67.pep																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
--	-----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

20	1	MPSETVGSIV	NVGVD	ESVGF	SPFP	PSIQHF	YRFH	RIHRIR	LFRPP	GPMQL			
	51	NRHSHGSGNL	GRGV	WATVLS	DKFPC	GQVRI	PACAG	MTNFE	IAVL	SGMTVR			
	101	VFYCARPAPV	NGGRL	KMPSE	GSDG	IGIGES	EAVA	HAQRGF	VGFE	AGVFQA			
	151	SPVVAVAGV	QQA	GRDVYA	HARH	RAEAQA	AAVA	FLIGV	FLRMS	VRINR			
	201	NCCVSI	TRVG	GKST	CYFFSR	IDAV	SDVSVG	DART	DIGFEF	VVEFE	IVNGG		
25	251	QAERRNGVEC	AVFL	MERLLV	FYVK	LVA	AKS	FIIL	SFQ	LFY	VHG	FIVV	PF
	301	PVTGI	IRGDA	PAAE	VVADR	H	PGVD	GMRTDV	SEII	AYRAYF	VFAWS	GWFR	I
	351	IVGNA	FGGVG	*									

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 30 Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

	1	ATGTTTGCTT	TTT	TAGAAGC	CTTT	TTTGTC	GAAT	ACGGTT	ATG	CGGCTGT
	51	TTTTTTTGTA	TTG	TCATCT	GCGG	TTTCGG	CGTG	CCGATT	CCC	GAGGATT
35	101	TGACCTTG	GT	AACAGG	CGGC	GTGAT	TTTCGG	GTAT	GGGTTA	TACCAATCCG
	151	CATATTAT	GT	TTGCAG	TCGG	TATG	CTCGC	GTAT	TGGTCG	GGGACGGCAT
	201	CATGTT	CGCC	GCCGG	ACGAA	TTTGG	GGGCA	GA	ArAr	TCCTA
	251	CTATTG	CGsG	CATCAT	GACG	CCG	AAACGTT	ATGAG	CAGT	TCAGGAAAAA
	301	TTCGAC	AAAT	ACGGT	AACTG	GGTCT	TATTT	GTCG	CCCCGTT	TCCTGCCCCG
	351	TTTGAGA	ACG	GCCGT	TATTTG	TTACAG	CCCG	TATCAG	CCCG	AAGGTTTCAT
40	401	ACTTGCG	TTT	TATCAT	TATG	GATG	GA	CTGG	CCGCA...	

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

	1	MFAFLEAFFV	EYGYAA	VFFV	LVICG	FGVPI	PEDL	TLVTGG	VISGM	GYTNP
	51	HIMFAVG	MGLG	VLVGD	GIMFA	AGRIW	GQXXL	XFXPI	AXIMT	PXRYEQVQEK
	101	FDKYGN	WVLF	VARFL	PGLRT	AVFVT	AGISR	KVSYL	RFIIM	DGLAA...

### 45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

	1	ATGTTTGCTT	TTT	TAGAAGC	CTTT	TTTGTC	GAAT	ACGGTT	ATG	CGGCTGT
	51	TTTTTTTGTA	TTG	TCATCT	GCGG	TTTCGG	CGTG	CCGATT	CCC	GAGGATT
	101	TGACCTTG	GT	AACAGG	CGGC	GTGAT	TTTCGG	GTAT	GGGTTA	TACCAATCCG
50	151	CATATTAT	GT	TTGCAG	TCGG	TATG	CTCGC	GTAT	TGGTCG	GGGACGGCAT
	201	CATGTT	CGCC	GCCGG	ACGAA	TTTGG	GGGCA	GAAA	ATCCTA	AGGTTCAAAC
	251	CTATTG	CGCG	CATCAT	GACG	CCGAA	ACGTT	ATGAG	CAGT	TCAGGAAAAA
	301	TTCGAC	AAAT	ACGGT	AACTG	GGTCT	TATTT	GTCG	CCCCGTT	TCCTGCCCCG
	351	TTTGAGA	ACG	GCCGT	TATTTG	TTACAG	CCCG	TATCAG	CCCG	AAGGTTTCAT
	401	ACTTGCG	TTT	TATCAT	TATG	GATG	GA	CTGG	CCGCA	CTGAT
55	451	ATTTG	GATTT	ATCTG	GGCGA	ATACG	GTGCG	CACAA	CATCG	ATTGGCTGAT

-410-

501 GGCAGAAATG CACAGCCTGC AATCGGGTAT TTTTGTATC TTGGGTATAG  
 551 GTGCGACCGT TGTGCTTGG ATTTGGTGA AAAACGCCA ACGTATCCAG  
 601 TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA  
 651 GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP  
 51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL RFKPIARIMT PKRYEQVQEK  
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP  
 151 IWIYLGEYGA HNIDWLMAMK HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ  
 201 FYRSKLKEKR AQRKAATAK KAAQSKQ\*

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H. influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

15 Orf78: 4 FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61  
 FL FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GV  
 DedA: 20 FLIGFFTEYGYWAVLFLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGV 79  
 20 Orf78: 62 LVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121  
 L GD M+ GRI+G L F PI I+T R V+EKF +YGN VLFVARFLPGLR  
 DedA: 80 LAGDSCMYWLGRIYGTKILRFRPIRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139  
 Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145  
 +++ +GI+R+VSY+RF+++D AA  
 25 DedA: 140 IYMVSGITRRVSYRVFLIDFCAA 163

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N. meningitidis*:

30 orf78.pep 10 20 30 40 50 60  
 MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG  
 orf78a MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG  
 35 10 20 30 40 50 60  
 orf78.pep 70 80 90 100 110 120  
 VLVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRT  
 orf78a VLVGDGIMFAAGRIWGQKILFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT  
 40 70 80 90 100 110 120  
 orf78.pep 130 140  
 AVFVTAGISRKVSYLRFIIMDGLAA  
 orf78a AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSQSGIFIA  
 45 130 140 150 160 170 180

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

1 ATGTTTGCCC TTTTGAAGC CTTTTTTGTC GAATACGGCT ATGCGGCCGT  
 51 GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCGAGGATT  
 101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG  
 151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT  
 201 CATGTTTCGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC  
 251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA  
 301 TTCGACAAAT ACGGCAACTG GGTGTTATTT GTCGCTCGTT TCCTGCCCCG  
 351 TTTGCGGACT GCCGTTTTCG TTACCGCCCG CATCAGCCGC AAAGTATCGT  
 401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC  
 451 GTTTGGATT ACTTGGCGCA GTACGGCGCG CACAACATCG ATTGGCTGAT



-411-

```

501 GGCAGAAATG CACAGCCTGC AATCCGGCAT CTTTCATCGCA TTGGGCGTGC
551 TGGCGCGCGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

```

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

```

1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMYTNP
51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFILM DGLAALISVP
151 VWIYLGEYGA HNIDWLMAMK HSLQSGIFIA LGVLAALAW FWRKRHHYQ
10 201 LYRAQLSEKR AKRKAEEKAA KAAQKQQ*

```

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

```

15 orf78a.pep      10      20      30      40      50      60
      MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMYTNP HIMFAVGMLG
      |||:|||||||||||||||||||||||||||||||||||||||||||||||||
      orf78-1      10      20      30      40      50      60
      MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMYTNP HIMFAVGMLG

20 orf78a.pep      70      80      90      100     110     120
      VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
      |||||||||||||||||||:|||||||||||||||||
      orf78-1      70      80      90      100     110     120
      VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT

25 orf78a.pep      130     140     150     160     170     180
      AVFVTAGISRKVSYLRFIMDGLAALISVPVWIYLGEYGAHNIDWLMAMK HSLQSGIFIA
      |||:|||||||||:|||||||||:|||||||||:|||||||||:|||||
      orf78-1      130     140     150     160     170     180
      AVFVTAGISRKVSYLRFIIMDGLAALISVPIWIYLGEYGAHNIDWLMAMK HSLQSGIFVI

30 orf78a.pep      190     200     210     220
      LGVLAALAWFWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQQX
      ||: ||::||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
      orf78-1      190     200     210     220
      LGIGATVVAWIWWKQRQIQFYRSKLKEKRAKRAKAAKAAQSKQX

35 orf78a.pep      190     200     210     220

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

```

40 orf78.pep      XXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF 137
      |||||||||||||||||||
      orf78ng      YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF 32

      orf78.pep      IIMDGLAA 145
      :|||||
45 orf78ng      LIMDGLAALISVPVWIYLGEYGAHNIDWLMAMK HSLQSGIFIALGVLAALAWFWRKR 92

```

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

```

1 ..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWIYL
51 GEYGAHNIDW LMAKMHSLS GIFIALGVLA AALAWFWRK RRHYQLYRAQ
101 LSEKRAKRA EKAAKAAQK QQ*

```

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

```

1 atgttttgcct tttTggaagc CTTTTTTGTC GAAatcggCt atgcGGCCGT
51 GTTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT
101 TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
151 CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT
201 GATGTTTGCC GCCGGACGCA TCTGGGGGCA GAAATCCTC AAGTTCAAAC
251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA
301 TTCGACAAAT ACGGCAACTG GGTCTGTTT GTCGCCCGTT TCCTGCCGGG

```

-412-

5  
351 TTTGCGGACT GCCGTTTTTCG TTACCGCCCG CATCAGCCGC AAAGTATCGT  
401 ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCC  
451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT  
501 GGCGAAAATG CACAGCCTGC AATCGGGCAT CTCATCGCA TTGGGCGTGC  
551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG  
601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA  
651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

10  
1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP  
2 HMFVAVGMLG VLAGDGVMA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK  
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRLIM DGLAALISVP  
151 VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAALAW FWWRRRHHYQ  
201 LYRAQLSEKR AKRKAEEKAAK KAAQKQQ\*

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15  
10 20 30 40 50 60  
orf78-1.pep MFALFEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHMFVAVGMLG  
:::||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
orf78ng-1 MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHMFVAVGMLG  
10 20 30 40 50 60  
20  
70 80 90 100 110 120  
orf78-1.pep VLVDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT  
:::||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
orf78ng-1 VLAGDGVMAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT  
25  
70 80 90 100 110 120  
30  
130 140 150 160 170 180  
orf78-1.pep AVFVTAGISRKVSYLRLIMDGLAALISVPIWIYLGEYGAHNIDWLMAKMHSLSGIFIA  
||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
orf78ng-1 AVFVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSGIFIA  
130 140 150 160 170 180  
35  
190 200 210 220  
orf78-1.pep LGIGATVVAWIWWKKRQRIQFYRSKLKEKRAQRKAAKAAQSKQX  
||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
orf78ng-1 LGVLAALAWFWWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQQX  
190 200 210 220

Furthermore, orf78ng-1 shows homology to the dedA protein from *H.influenzae*:

40 sp|P45280|YG29\_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir||D64133 dedA  
protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)  
>gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212  
Score = 223 bits (563), Expect = 7e-58  
Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)  
45 Query: 5 LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHMFVAVGMLGVL 62  
L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GVL  
Sbjct: 21 LIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGVL 80  
50 Query: 63 AGDGVMAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122  
AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +  
Sbjct: 81 AGDSCMYWLGRIYGTILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140  
55 Query: 123 FVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSGIFIALG 182  
++ +GI+R+VS+Y+RF+++D AA+ISVP+WIYLG E GA N+DWL ++ Q I+I +G  
Sbjct: 141 YMVSGITRRVS+YR+VFLIDFCAAII+SVPIWIYLGELGAKNLDWLHTQIQKGQIVIIYIFIG 200  
Query: 183 VL 184  
L  
Sbjct: 201 YL 202  
60

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 87

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```

1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
10 201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAAATTCCC GTTACCCTGA AATTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA C...
```

- 15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK QLKKEGDKIP VTLKEFNAKA QTVQLEVKIA PMPAMNH...
```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

```

20 1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
25 201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAAATTCCC GTTACCCTGA AATTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA CGGTATCAC
451 CACGCGGAAG CGCATCAGCA CTAA
```

- 30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

```

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK QLKKEGDKIP VTLKEFNAKA QTVQLEVKIA PMPAMNHGHH
151 HGEAHQH*
```

- 35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N.meningitidis*:

```

40      10      20      30      40      50      60
orf79.pep MKKLLAAVMMAGLAGAVSAAAGVHVEDGWAR TTVEGMKIGGAFMKIHNDEAKQDFLLGGSS
      |||||||:|||||:|||||:|||||:|||||:|||||
orf79a    MKKLLAAVMMAGLAGAVSAAAGIHVEDGWAR TTVEGMKMGGA FMKIHNDEAKQDFLLGGSS
      10      20      30      40      50      60

45      70      80      90      100     110     120
orf79.pep PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
      |||||||:|||||:|||||:|||||:|||||:|||||
orf79a    PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
50      70      80      90      100     110     120
```

```

                                130      140
orf79.pep  VTLKFKNAKAQTQVQLEVKIAPMPAMNH
            |||||
5 orf79a    VTLKFKNAKAQTQVQLEVKTAPMSAMDGHGHHGEAHQHXX
                                130      140      150

```

	1	ATGAAANAAC	TATTGCGAGC	CGTGATGATG	GCAGGTTTGG	CAGGCGCGGT
10	51	TTCCGCCGCC	GGAATCCACG	TTGAGGACGG	CTGGGCGCGC	ACCACCGTCG
	101	AAGGTATGAA	AATGGGCGGC	GCGTTCATGA	AAATCCACAA	CGACGGAAGCC
	151	AAACAAGACT	TTTTGCTCGG	CGGAAGCAGC	CCTGTTGCCG	ACCGCGTCGA
	201	AGTGCATACC	CATATCAATG	ATAACGGTGT	GATGCGGATG	CGCGAAGCTCG
	251	AAGGCGGCGT	GCCTTTGGAG	GCGAAATCCG	TTACCGAACT	CAAACCCGGC
15	301	AGCTATCATG	TCATGTTTAT	GGGTNTGAAA	AAACAATTAA	AAGANGGCGA
	351	CAAGATTCCC	GTTACCTCGA	AATTTAAAAA	CGCCAAGCA	CAAACCGTCC
	401	AACTGGAAGT	CAAAACCGCG	CCGATGTCGG	CAATGGACCA	CGGTTCATCAC
	451	CACGGCGAAG	CGCATAGCA	CTAA		

20

1	<u>MXLLAAVMM</u>	<u>AGLAGAVSAA</u>	<u>GIHVEDGWAR</u>	<u>TTVEGMMKMG</u>	<u>AFMKIHNDEA</u>
51	<u>KQDFLLGGSS</u>	<u>PVADRVVHT</u>	<u>HINDNGVMRM</u>	<u>REVEGGVPLE</u>	<u>AKSVTELPKG</u>
101	<u>SYHVMFMGXX</u>	<u>KQLKXGDKIP</u>	<u>VTLKFKNAKA</u>	<u>QTVQLEVKTA</u>	<u>PMSAMDHGHH</u>
151	HGEAHOH*				

			10	20	30	40	50	60
25	orf79a.pep		MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGAFMKIHNDEAKQDFLLGGSS					
	orf79-1		MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGAFMKIHNDEAKQDFLLGGSS					
			10	20	30	40	50	60
30	orf79a.pep		70	80	90	100	110	120
			PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELPKGSYHVMFMGXXXQLKXGDKIP					
	orf79-1		PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELPKGSYHVMFMLKKQLKEGDKIP					
			70	80	90	100	110	120
35								
	orf79a.pep		130	140	150			
			VTLKFKNAKAQTVQLEVKTAPMSAMDGHGHHGGEAHQHXX					
	orf79-1		VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHGGEAHQHXX					
40			130	140	150			

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

10

15

20

25

30

35

40

45

50

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

**Example 88**

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

```

5      1  ATGACGGTAA CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
      51  ATATCTGATT ACGGGCATT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
     101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
     151  CCGAAGCAAT GGC GGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
     201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAAACCGGA TTGTTTGCCG
     251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
    301  CGGATTCCGG TTGTGAAATC CATCTATTCT AGTGTGAAAA AAGTATCCGA
    351  ATAcgTGCTG TCCGACAGCA GCCGTTCTGT TAAACGCCC GTACTCGTGC
    401  CGTTTCCCCA GCCCGGTATT TGGACGATyG CTTTCGTGTC AGGGCAGGTG
    451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAs GACGGCGATT ATCTTCCGT
    501  GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTACTAT ATTATGGTAA
    551  AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AsCATTGAAA
    601  TATGTGATTT CGCTGGGTAT GGTATCCCT GACGACCTGC CCGTCAAAAC
    651  ATTGGCAsGA CCTATGCCGT CTGAAAAGGC GGATTGCCC GAACAACAAT
    701  AA

```

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

```

20      1  MTVTAAEGGK AAKALKKYLI TGILVWLPiA VTVWVVSyIV SASDQLVNLL
      51  PKQWRPQYVL GFNIPGLGVI VAIaVLFVTG LFAANVLGRQ ILAAWDSLLG
     101  RIPVVKSIYS SVKKVSEYVL SDSSRSFKTP VLVFPFPQPGI WTIAFVSGQV
     151  SNAVKAALPX DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEXLK
     201  YVISLGMVIP DDLPVKTLAX PMPSEKADLP EQQ*

```

25 Further work revealed the complete nucleotide sequence <SEQ ID 741>:

```

30      1  ATGACGGAAC nTGC GGCCGA AGGCGGCAAA GCTGCCAArG CGTTAAAAAA
      51  ATATCTGATT ACGGGCATT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
     101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
     151  CCGAAGCAAT GGC GGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
     201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAAACCGGA TTGTTTGCCG
     251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
     301  CGGATTCCGG TTGTGAAATC CATCTATTCT AGTGTGAAAA AAGTATCCGA
     351  ATCGCTGCTG TCCGACAGCA GCCGTTCTGT TAAACGCCC GTACTCGTGC
     401  CGTTTCCCCA GCCCGGTATT TGGACGATTG CTTTCGTGTC AGGGCAGGTG
     451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAG GACGGCGATT ATCTTCCGT
     501  GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTACTAT ATTATGGTAA
     551  AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AGCATTGAAA
     601  TATGTGATTT CGCTGGGTAT GGTATCCCT GACGACCTGC CCGTCAAAAC
     651  ATTGGCAGGA CCTATGCCGT CTGAAAAGGC GGATTGCCC GAACAACAAT
     701  AA

```

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

```

45      1  MTEXAAEGGK AAKALKKYLI TGILVWLPiA VTVWVVSyIV SASDQLVNLL
      51  PKQWRPQYVL GFNIPGLGVI VAIaVLFVTG LFAANVLGRQ ILAAWDSLLG
     101  RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPFPQPGI WTIAFVSGQV
     151  SNAVKAALPK DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
     201  YVISLGMVIP DDLPVKTLAG PMPSEKADLP EQQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N. meningitidis*:

```

55      10      20      30      40      50      60
orf98.pep  MTVTAAEGGKA AAKALKKYLI TGILVWLPiA VTVWVVSyIV SASDQLVNLL PKQWRPQYVL
           ||  |||||
orf98a     MTEPAAEGGKA AAKALKKYLI TGILVWLPiA VTVWVVSyIV SASDQLVNLL PKQWRPQYVL
           10      20      30      40      50      60

```

-417-

		70	80	90	100	110	120
	orf98.pep	GFNIPGLGVIVAI	AVLFVTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKKVSEYVL			
5	orf98a	GFNIPGLGVIVAI	AVLFVTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKKVXSLL			
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf98.pep	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSVYVPTTPNPTGGYY			
	orf98a	SDSSRSFKTPVLV	PFPPQSGIWTIAFVSGQV	SNVKAALPKDGDYLSVYVPTTPNPTGGYY			
		130	140	150	160	170	180
15	orf98.pep	IMVKKSDVRELDMSV	DEXLKYVISLGMVIPDDL	PVKTLAXPMPSEKADLPEQQX			
	orf98a	IMVKKSDVRELDMSV	DEALKYVISLGMVIPDDL	PVKTLAGPMPSEKADLPEQQX			
		190	200	210	220	230	
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
25	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG
	251	CAAACGTATT	GGGCCGGCAG	ATTCTTGCCC	CGTGGGACAG	CTTGTTGGGG
	301	CGGATTCCGG	TTGTGAAGTC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAAACACCA	GTA CTCTGTC
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCTGTGC	CGGTACAGTG
30	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTTCCGT
	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTTGCCC	GAACAACAAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFVTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKSXSLL	SDSSRSFKTP	VLVPPQSGI	WTIAFVSGQV
40	151	SNVKAALPK	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK
	201	YVISLGMVIP	DDL PVKTLAG	PMPSEKADLP	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
	orf98a.pep	MTEPAAEGGKA	AKALKKYLIT	GI	LVLPIAVT	VWVSIVSASD	QLVNLLPKQWRPQYVL
45	orf98-1	MTEXAAEGGKA	AKALKKYLIT	GI	LVLPIAVT	VWVSIVSASD	QLVNLLPKQWRPQYVL
		10	20	30	40	50	60
		70	80	90	100	110	120
50	orf98a.pep	GFNIPGLGVIVAI	AVLFVTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKKVXSLL			
	orf98-1	GFNIPGLGVIVAI	AVLFVTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKKVSESLL			
		70	80	90	100	110	120
55	orf98a.pep	SDSSRSFKTPVLV	PFPPQSGIWTIAFVSGQV	SNVKAALPKDGDYLSVYVPTTPNPTGGYY			
	orf98-1	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSVYVPTTPNPTGGYY			
		130	140	150	160	170	180
60	orf98a.pep	IMVKKSDVRELDMSV	DEALKYVISLGMVIPDDL	PVKTLAGPMPSEKADLPEQQX			
	orf98-1	IMVKKSDVRELDMSV	DEALKYVISLGMVIPDDL	PVKTLAGPMPSEKADLPEQQX			
65		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

5	orf98.pep	MTVTAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
	orf98ng	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
10	orf98.pep	GFNIPGLGVIVAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL	120
	orf98ng	GFNIPGLGVIVAVLFTGLFAANVLGRQILAAWDSLLXRPVVKSIYSSVKKVSESL	120
15	orf98.pep	SDSSRSFKTPVLVPPFQPGIWTIAFVSGQVSNVKAALPXDGDYLSVYVPTTNPNTGGYY	180
	orf98ng	SDSSRSFKTPVLVPPFQSGIWTIAFVSGQVSNVKAALPQDGDYLSVYVPTTNPNTGGYY	180
	orf98.pep	IMVKKSDVRELDMSVDEXLKYVISLGMVIPDDLVPKTLAXPMPSEKADLPEQQ	233
	orf98ng	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPPEKAELPEQQ	233

- 20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

25	1	MTEPAAEGGK	AAKALKKYL	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLX
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPPFQSGI	WTIAFVSGQV
	151	SNAVKAALPQ	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK
	201	YVISLGMVIP	DDLVPKTLAG	PMPPEKAELP	EQQ*	

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

30	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACAGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ACCAGCTTGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCCGGGCT
	201	CGGCGTTATT	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG
35	251	CAAACGTGTT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CCTGTTgggg
	301	cggattCCGG	TTGTCAAATC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAAACGCCG	GTACTCGTGC
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCTGTGC	CGGTCAGGTG
	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGCG	GATGGCGATT	ATCTTTCCGT
40	501	GTATGTCCCG	ACCACGCCCA	ACCCGACCGG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAC
	651	ATTGGCAGGA	CCTATGCCGC	CTGAAAAGGC	GGAGTTGCCC	GAACAACAAT
	701	AA				

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

45	1	MTEPAAEGGK	AAKALKKYL	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPPFQSGI	WTIAFVSGQV
	151	SNAVKAALPQ	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK
	201	YVISLGMVIP	DDLVPKTLAG	PMPPEKAELP	EQQ*	

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

50	orf98-1.pep	MTEXAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
	orf98ng-1	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
55	orf98-1.pep	GFNIPGLGVIVAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	120
	orf98ng-1	GFNIPGLGVIVAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	120



-419-

orf98ng-1		GFNIPGLGVIVAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	70	80	90	100	110	120
orf98-1.pep		SDSSRSFKTPVLVFPFPQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
orf98ng-1		SDSSRSFKTPVLVFPFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
orf98-1.pep		IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLPVKTLAGPMPSEKADLPEQQX	190	200	210	220	230	
orf98ng-1		IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLPVKTLAGPMPPEKAELPEQQX	190	200	210	220	230	

- Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 89

- The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

1	ATgAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	G <sub>s</sub> GgTACTCA
201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGChCGTAA	AGGCCkCAAG
251	ssCGsGCTTG	CCTTGAACAA	GGCGGGTTTG	GCGTATTTTG	AAGGGCGTTT
301	TGAAAAGGCG	GAAC TAGAAG	CCTCACGCGT	GTTGGTCAAC	AAAGTAGGCC
351	GaGAGACAAC	CGGACTTTGG	CATTGATGCT	GrCGCGCAC	GCCGCCGGAC
401	AGATGGAAAA	CATCGAsTG	CGCGACCGTT	ATCTTGCGGA	AATCGCCAAA
451	CTGCCGGA	AACAGCAGCT	TTCCCGTTAT	CTTTTGTG	CGGAATCGGC
501	GTTGAACCGG	CGCGATTACG	AAGCGGCGGA	AGCCAATCTT	CATGCGGCGG
551	CGAAGATGAA	TGCCAACCTT	ACGCGCCTCG	TGCGTCTGCA	.ATTGCTTAC
601	GCTTTCGACA	GGGGCGACGC	GTTGCAGGTT	CTGGCAAAAA	CCGAAAAACT
651	TTCCAAGGCG	GGCGGTTTGG	GCAATCGGA	AATGGAACGG	TATCAAAATT
701	GGGCATATCC	GTCGCCAGCT	GGCGGATGCT	GCCGATGCCG	CCGCTTTGAA
751	AACCTGCCTG	AAGCGGATTC	CCGACAGCCT	CAAAAACGGG	GAATTGAGCG
801	TATCGGTTGC	GGAAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATGCGGTC
851	AAATGGGTCA	AACAGCATT	TCCGCA <sub>s</sub> AAC	CGCCGCCCGG	AGCTTTTGGA
901	AGCCTTTGTC	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
951	CCATCGATTT	TGCCGATGCT	TGGCTGAAAG	AACAGCCCGA	TAACGCGCTT
1001	CTGCTGATGT	ATCTCGGTCG	GCTCGCCTTC	GGCCGCAAA	TTTGGGGCAA
1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATTGC	ATTAAAGCCG	AGTATTTC
1101	CGCGTTTGGT	TCTAACAAAG	GTTTTCGACG	AAATCGGAGA	ACCGCAGAAG
1151	GCGGAGGCGC	AC...			

- This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

1	MKTVVWIVVL	FAAVGLALA	SGIYTGdVYI	VLGQTMlRIN	LHAFVLGSLI
51	AVVVWYFLFK	FIIGVLNIPE	KMQRFGSARK	GXKXXLALNK	AGLAYFEGRF
101	EKAELASRV	LVNKVGRDNR	TLALMLXAHA	AGQMenIXXR	DRYLAEIAKL
151	PEKQQLSRYL	LLAESALNRR	DYEA <sub>EA</sub> NLH	AAAKM <sub>N</sub> ANLT	RLVRLXIRYA
201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQLA	DAADAALKT
251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWKQHYP	XNRRPELLEA
301	FVESVRELGE	REQQKAIDFA	DAWLKEQPDN	ALLLMYLGRl	AFGRKLWGKA
351	KGYLEASIAL	KPSISARLVL	TKVFDEIGEP	QKAEAH...	

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GCGTACTCAA

-420-

5  
10  
15  
20

```

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GCGCGGTAAA GGCCGCAAGG
251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT
301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCGCG
351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA
401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG
451 CCGGAAAAC AGCAGCTTTC CCGTTATCTT TTGTGGCGG AATCGGCGTT
501 GAACCGGCGC GATTACGAAG CCGCGGAAGC CAATCTTCAT GCGGCGGCGA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT
601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAATTTTC
651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTTGGG
701 CATACCGCCG CCAGCTGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCCTGAAGC GGATTCCCGA CAGCCTCAAA AACGGGGAAT TGAGCGTATC
801 GGTTGCGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCCGAGCT TTTGGAAGCC
901 TTTGTCGAAA GCGTGCCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
951 CGATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC
1001 TGATGTATCT CGGTCGGCTC GCCTACGGCC GCAAACCTTG GGGCAAGGCA
1051 AAAGGCTACC TTGAAGCGAG CATTGCAATTA AAGCCGAGTA TTTCCGCGCG
1101 TTTGGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACCG CAGAAGGCGG
1151 AGGCGCAGCG CAACTTGGTT TTGGAAGCCG TCTCCGATGA CGAACGTCAC
1201 GCAGCGTTAG AGCAGCATAG CTGA

```

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

25  
30

```

1 MKTVVWIVVL FAAAVGLALA SGIYTGVDYI VLQOTMLRIN LHAFVLGSLI
51 AVVVWYFLFK FIIGVLNIPE KMORFGSARK GRKAALALNK AGLAYFEGRF
101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAETAKL
151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNNANLT RLVRLQLRYA
201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT
251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP HNNRRPELLEA
301 FVESVRFLGE REQQKAIDFA DAWLKEQPDN ALLMYLGLRL AYGRKLWGKA
351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRNLV LEAVSDDERH
401 AALEQHS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

35 *meningitidis*:

40  
45  
50  
55  
60

```

              10      20      30      40      50      60
orf100.pep  MKTVVWIVVLFAAAVGLALASGIYTGVDYIIVLGOTMLRINLHAFVLGSLIAVVVWYFLFK
              |||||
orf100a     MKTVVWIVVLFAAAXGLALASGIXTGVDYIIVLGOTMLRINLHAFVLGSLIAVVVWYFLFK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf100.pep  FIIGVLNIPEKMQRFGSARKGXXXXLALNKAGLAYFEGRFKAELEASRVLVNKGVRDNR
              |||||
orf100a     FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFKAELEASRVLVNKEAGDNR
              70      80      90      100     110     120

              130     140     150     160     170     180
orf100.pep  TLALMLXAHAAAGQMENIXRDRYLAETAKLPEKQQLSRYLLLAESALNRRDYAAEANLH
              |||||
orf100a     TLALMLGAHAAGQMENIELRDRYLAETAKLPEKQQLSRYLLLAESALNRRDYAAEANLH
              130     140     150     160     170     180

              190     200     210     220     230     240
orf100.pep  AAKMNNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA
              |||||
orf100a     AAKMNNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX
              190     200     210     220     230     240

              250     260     270     280     290     300
orf100.pep  DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPXNNRRPELLEA
              |||||
orf100a     DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNNRRPELLEA
              250     260     270     280     290     300

```

-421-

		310	320	330	340	350	360
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRLA	FGRKLWGKAKGYLEASIAL				
5	orf100a	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLXYLGRLAY	GRKLWGKAKGYLEASIAL				
		310	320	330	340	350	360
		370	380				
10	orf100.pep	KPSISARLVLT	KVFDEIGEPQKAEAH				
	orf100a	KPSISARLVLA	KVFDETGEPQKAEAQ	RNLVLASVAEENRPSAETHX			
		370	380	390	400		

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCTCG	TTTGCCGCCG	CNNTCGGGCT
51	GGCATTGGCG	TCGGGCATTN	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTAGG	TTGCGTGATT
151	GCCGTCGTGG	TGTGGTATTT	CCTGTTCAAA	TTCATCATCG	GCGTACTCAA
201	TANCCCCGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAAA	GGCCGCAAGG
251	CCGCGCTTGC	TTTGAACAAG	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTTT
301	GAAAAGGCGG	AACTTGAAGC	CTCGCGCGTA	TTGGGAAACA	AAGAGGCGGG
351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCCGGGCAGA
401	TGGAAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAAACTG
451	CCGAAAAGC	AGCAGCTTTC	CCGTTATCTT	TTGTTGGCGG	AATCGGCGTT
501	GAACCGGCGC	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCGA
551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
601	TTGACACGGG	GCGACGCGTT	GCAGGTTCTG	GCAAAAACCG	AAAAANTTTC
651	CAAGGCGGGC	GCGTNGGGCA	AATCGGAAAT	GGAACGGTAT	CAAAATTGGG
701	CATACCGCCG	CCAGCTGNCG	GATGCTGCCG	ATGCCGCGCG	TTTGAAAACC
751	TGCCTGAAGC	GGATTCCCGA	CAGCCTCAAA	AACGGGGAAT	TGAGCGTATC
801	GGTTGCGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GACCCGAACT	TTTGGAAGCN
901	TTTGTCGAAA	GCGTGCGCTT	TTTGGGCGAA	CGCGATCAGC	AGAAAGCCAT
951	CGATTTTGCC	GATGCTTGGC	TGAAAGAACA	GCCCGATAAT	GCGCTTCTGC
1001	TGANGTATCT	CGGTGCGCTC	GCCTACGGCC	GCAAACTTTG	GGGCAAGGCA
1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCCGAGTA	TTTCCGCGCG
1101	TTTGGTTCTG	GCAAAGGTTT	TTGACGAAAC	CGGAGAACCG	CAGAAGGCGG
1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGAGGA	AAACCGNCT
1201	TCCGCCGAAA	CCATTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

1	MKTVVWIVVL	FAAAXGLALA	SGIXTGDVYI	VLGQTMLRIN	LHAFVLGSLI
51	AVVVWYFLFK	FIIGVLNXPE	KMQRFGSARK	GRKAALALNK	AGLAYFEGRF
101	EKAEELEASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLQLRYA
201	FRGDALQVL	AKTEKXSKAG	AXGKSEMERY	QNWAYRRQLX	DAADAALKT
251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWVKQHYP	HNRRPELLEA
301	FVESVRFLGE	RDQQAIDFA	DAWLKEQPDN	ALLLXYLGR	LAAYGRKLWGKA
351	KGYLEASIAL	KPSISARLV	LAKVFDETGEP	QKAEARNLV	LASVAEENRP
401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

50	orf100a.pep	10	20	30	40	50	60
	orf100-1	10	20	30	40	50	60
55	orf100a.pep	70	80	90	100	110	120
	orf100-1	70	80	90	100	110	120
60	orf100a.pep	130	140	150	160	170	180
	orf100-1	130	140	150	160	170	180

-422-

		130	140	150	160	170	180
5	orf100a.pep	190	200	210	220	230	240
	orf100-1	190	200	210	220	230	240
10	orf100a.pep	250	260	270	280	290	300
	orf100-1	250	260	270	280	290	300
15	orf100a.pep	310	320	330	340	350	360
	orf100-1	310	320	330	340	350	360
20	orf100a.pep	370	380	390	400		
	orf100-1	370	380	390	400		

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from

*N.gonorrhoeae*:

30	orf100.pep	MKTVVWIVVLF	AAA	VGLALASGIY	TGDVYIVL	GGTMLRINLHAF	VLGSLIAVVVWY	FLFK	60
	orf100ng	MKTVVWIVVLF	AAA	VGLALASGIY	TGDVYIVL	GGTMLRINLHAF	VLGSLIAVVVWY	FLFK	60
35	orf100.pep	FIIGVLNIP	EKMQRFGS	ARKGXKXXL	ALNKA	LAYFEG	RFEKAELEAS	RVLVNKVGRDNR	120
	orf100ng	FIIGVLNIP	ENMRRSGS	ARKGRKAAL	ALNKA	LAYFEG	RFEKAELEAS	RVLVNKEAGDNR	120
40	orf100.pep	TLALMLX	AHAAGQ	MENIXRDR	YLAEIAKL	PEKQQLS	RYLLLAES	ALNRRDYEA	180
	orf100ng	TLALMLG	AHAAGQ	MENIELRDR	YLAEIAKL	PEKQQLS	RYLLLAES	ALNRRDYEA	180
45	orf100.pep	AAAKMNANL	TRLVRLX	IRYAFDRG	DALQVLAK	TEKLSKAG	ALGKSEMER	YQNWAYRRQLA	240
	orf100ng	AAAKMNANL	TRLVRLQ	RLYAFDRG	DALQVLAK	TEKLSKAG	ALGKSEMER	YQNWAYRRQMA	240
50	orf100.pep	DAADAAALK	TCLKRIP	DSLKNGEL	SVSVAEK	YERLGLY	ADAVKWKQ	HYPNRRPELLEA	300
	orf100ng	DAADAAALK	TCLKRIP	DSLKNGEL	SVSVAEK	YERLGLY	ADAVKWKQ	HYPHNRRPELLEA	300
55	orf100.pep	FVESVRFL	GEREQQA	IDFADAWL	KEQPDN	ALLMYLGR	LAFGRKLW	GKAKGYLEASIAL	360
	orf100ng	FVESVRFL	GEREQQA	IDFADSWL	KEQPDN	ALLMYLGR	LAYGRKLW	GKAKGYLEASIAL	360
	orf100.pep	KPSISARLV	LTKVFEI	GEPEKAE	AH				386
	orf100ng	KPSIPARLV	LAKVFDE	TASQSKAE	AQRNLV	LASVAGEN	NRPSAETR		405

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

	1	ATGAAAACGG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTCGGACT	
60	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTA	CTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TT	CGTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTTAA	TTCA	TATCATCG	GCGTACTCAA
	201	TATCCCGGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCCGCAAGG	
	251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTCGA	AGGGCGTTTT	
	301	GAAAAGGCGG	AACTCGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAGCCCGG	
65	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGGACAGA	
	401	TGGAAAATAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAAACTG	

-423-

5  
10  
15

```

451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGGCGTT
501 AAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC
601 TTCGATCGGG GCGATGCGTT GCAGGTTCTG GCAAAAaccG AAAAACTTTC
651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
701 CATACGCGCG CCAGATGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCCTGAAGC GGATTCCCGA CAGCCTCAAA AACGGGGAAT TGagcGTATC
801 GGTTCGCGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
901 TTTGTCGAAA GCGTGCCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
951 CGATTTTGCC GATTCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC
1001 TGATGTATCT CGGCCGGCTC GCCTACGGCC GCAAACCTTG GGGTAAGGCA
1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCCGGCGCG
1101 TTTGGTGTG GCAAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCCG
1151 AAGCACAGCG CAACTTGGTT TTGGCAAGCG TTGCCGGGGA AAACCGCCCT
1201 TCCGCCGAAA CCCGTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 756>:

20  
25

```

1 MKTVVWIVVL FAAAVGLALA SGIYTGdVYI VLGQTMLRIN LHAFVLGSLI
51 AVVVWYFLFK FIIGVNIPE NMRRSGSARK GRKAALALNK AGLAYFEGRF
101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
151 PEKQQLSRYL LLAESALNRR DYEAEEANLH AAAMNANLT RLVRLQLRYA
201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAAALKT
251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH HNRRPELLEA
301 FVESVRLGE REQQKAIDFA DSWLKEQPDN ALLLMYLGR LAYGRKLWGKA
351 KGYLEASIAL KPSIPARLV LAKVFEETAQS QKAEAQRLN LASVAGENRP
401 SAETR*

```

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

30  
35  
40  
45  
50  
55  
60  
65

```

      10      20      30      40      50      60
orf100-1.pep MKTVVWIVVLFAAAVGLALASGIYTGdVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
      |||||
orf100ng      MKTVVWIVVLFAAAVGLALASGIYTGdVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
      10      20      30      40      50      60

      70      80      90      100     110     120
orf100-1.pep FIIGVNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLVNKEAGDNR
      |||||
orf100ng      FIIGVNIPENMRRSGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLVNKEAGDNR
      70      80      90      100     110     120

      130     140     150     160     170     180
orf100-1.pep TLALMLGAHAAGQMENIELRDRLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH
      |||||
orf100ng      TLALMLGAHAAGQMENIELRDRLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH
      130     140     150     160     170     180

      190     200     210     220     230     240
orf100-1.pep AAAMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA
      |||||
orf100ng      AAAMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA
      190     200     210     220     230     240

      250     260     270     280     290     300
orf100-1.pep DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYHNNRRPELLEA
      |||||
orf100ng      DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYHNNRRPELLEA
      250     260     270     280     290     300

      310     320     330     340     350     360
orf100-1.pep FVESVRLGEREQQKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL
      |||||
orf100ng      FVESVRLGEREQQKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL
      310     320     330     340     350     360

      370     380     390     400
orf100-1.pep KPSISARLVLAKEVFEIGEPQKAEAQRLNVLAVSDDERHAALEQHSX
      ||||
orf100n      KPSIPARLVLAKEVFEETAQSQAQRLNVLAVSAGENRPSAETRX

```

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

10      1  ATGATGTTTT CTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTTCGTG
      51  GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTTCGTCAAT ATGGCGATGA
      101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
      151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
      201  CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
      251  ACGTCAAACG GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
      301  GCGGTGCTGC TCGCGCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
      351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
      401  TGTATsTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

20      1  MMFSWFKLFH LFFVISWFAG LFYLPRIFFN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVEGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
      101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYXVVEK PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

25      1  ATGATGTTTT CTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTTCGTG
      51  GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTTCGTCAAT ATGGCGATGA
      101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
      151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
      201  CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
      251  ACGTCAAACG GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
      301  GCGGTGCTGC TCGCGCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
      351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
      401  TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

35      1  MMFSWFKLFH LFFVISWFAG LFYLPRIFFN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVEGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
      101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVEFK PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40      orf102  3  FSWFKLFHLEFFVISWFAGLFYLPRIFFNMMAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
      HP1484  8  FLWVKAHFHVIAVISWMAALFYLPRLFVYHAENAHKKEFVGVVQIQEK--KLYSFIASPA 65

      orf102  63  GAVVEGAAIPFAAG---WWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWY 119
      HP1484  66  GFTLITGILMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRNARFY 125

      orf102  120 RVFNEIPXXXXXXXXXXXXXFKPF 142
      HP1484  126 RVFNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N. meningitidis*:

5		10	20	30	40	50	60
	orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVR	LYRFMSPL			
	orf102a	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVR	LYRFMSPL			
10		10	20	30	40	50	60
	orf102.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
	orf102a	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
15		70	80	90	100	110	120
	orf102.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
	orf102a	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
20		130	140				
	orf102.pep	VFNEIPVLLMVAALYXVVKPFX					
	orf102a	VFNEIPVLLMVAALYL	VVKPFX				
		130	140				

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

	1	ATGATGTTTT	CTTGGTTCAA	GCTGTTTCAC	TTGTTTTTTG	TCATTTCTGTG
	51	GTTTGCAGGG	CTGTTTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
25	101	TTGATGTGCC	GCGCGGCAAT	CCCGAGTATG	TGCGTCTGTC	GGGCATGGCG
	151	GTGCGGCTGT	ACCGTTTTAT	GTGCGCGTTG	GGCTTCGGCG	CGGTCGTGTT
	201	CGGCGCGGCG	ATACCGTTTG	CCGCCGGCTG	GTGGGGCAGC	GGCTGGGTAC
	251	ACGTCAAAC	TGTTTTGGGC	TTGATGCTCT	TGGCTTACCA	GTGTATTTGC
	301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
30	351	CTGGTACCGC	GTGTTCAACG	AAATCCCGT	GCTGCTGATG	GTGCGCGCGC
	401	TGTATCTGGT	CGTGTTCAAA	CCGTTTTGA		

This encodes a protein having amino acid sequence <SEQ ID 762>:

	1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVNMAMIDVPRGN	PEYVRLSGMA
	51	VRLYRFMSPL	GFGAVVFGAA	IPFAAGWWS	GWVHVKLCLG	LMLLAYQLYC
35	101	GVLRRRFQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYL	VVEK PF*

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

		10	20	30	40	50	60
	orf102a.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVR	LYRFMSPL			
40	orf102-1	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVR	LYRFMSPL			
		10	20	30	40	50	60
	orf102a.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
45	orf102-1	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
		70	80	90	100	110	120
	orf102a.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
	orf102-1	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYS	NAF	SHRWYR		
50		130	140				
	orf102a.pep	VFNEIPVLLMVAALYL	VVEKPF				
	orf102-1	VFNEIPVLLMVAALYL	VVEKPF				
		130	140				

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N. gonorrhoeae*:

-426-

```

    orf102.pep  MMFSWFKLFHLEFFVISWFAGLFYLPRIFFVNAMIDVPRGNPEYVRLSGMAVRLYRFMSPL  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    orf102ng    MMFSWFKLFHLEFFVISWFAGLFYLPRIFFVNAMIDAPRGNPEYVRLSGMAVRLYRFMSPL  60

5    orf102.pep  GFGAVVFGAAIPFAAGWGWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    orf102ng    GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR  120

10   orf102.pep  VFNEIPVLLMVAALYXVVFKEPF  142
               ||||||||||||||||||
    orf102ng    VFNEIPVLLMVAALYLVEFKPF  142

```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

15   1  ATGATGTTTT CTGGTTCAA GCTGTTTAC TTGTTTTTTG TCATTTCGTG
      51  GTTTGCAGGG CTGTTTACC TGCCGAGGAT TTTCTCAAT ATGGCGATGA
      101 TTGATGCGCC GCGCGGCAAT CCCGAGTATG TGCGCCTGTC GGGGATGGCG
      151 GTGCGGTTGT ACCGTTTTAT GTCGCCTTTG GGTTTCGGCG CGGTCGTGTT
      201 CCGCGCGGCG ATACCGTTTG CCGCgggccc GTGGGGCagc ggctggGTTT
      251 ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTATCA GTTGATTGTC
      301 GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
20   351 CTGGTACCGC GTGTTCACg aAATCCCCGT GCTGCTGATG GTTCCGCGCG
      401 TGTATCTGGT CGTGTTCAA CCGTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

25   1  MMFSWFKLFH LFFVISWFAG LFYLPRIFFN MAMIDAPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGRWGS GWVHVKLCLG LMLLAYQLYC
      101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVEFK PF*

```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

30   10      20      30      40      50      60
    orf102-1.pep MMFSWFKLFHLEFFVISWFAGLFYLPRIFFVNAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    orf102ng    MMFSWFKLFHLEFFVISWFAGLFYLPRIFFVNAMIDAPRGNPEYVRLSGMAVRLYRFMSPL
               10      20      30      40      50      60

35   70      80      90      100     110     120
    orf102-1.pep GFGAVVFGAAIPFAAGWGWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    orf102ng    GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
               70      80      90      100     110     120

40   130     140
    orf102-1.pep VFNEIPVLLMVAALYLVEFKPEX
               ||||||||||||||||||
    orf102ng    VFNEIPVLLMVAALYLVEFKPEX
               130     140

```

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```

45   gi|2314656 (AE000647) conserved hypothetical integral membrane protein
      [Helicobacter pylori] Length = 148
      Score = 79.2 bits (192), Expect = 1e-14
      Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

50   Query: 3  FSWFKLFHLEFFVISWFAGLFYLPRIFFVNAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62
      F W K FH+ VISW A LFYLPRIFFV A + V++ +LY F++
      Sbjct: 8  FLWVKAFHVIAVISWMAALFYLPRIFFVYHAENAHKKEFVGVVQIQEK--KLYSFIASPAM 65

55   Query: 63  GAVVFGAAIP-----FAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFS 115
      G + + F +G GW+H KL L ++LLAY YC +R + +
      Sbjct: 66  GFTLITGILMLLIEPTLFKSG----GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRN 121

      Query: 116 HRWYRVFNEIPXXXXXXXXXXXXFKPF 142
      R+YRVFNE P KPF
60   Sbjct: 122 ARFYRVFNEAPTILMILIVILVVVKPF 148

```



Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1  ATGGCAAAAA TGATGAAATG GCGGCTGTT GCGGCGGTCG CGGCGGCAGC
      51  GGTTTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
     101  TTACGGAAAC GGTACGGCGC GGC // .....
      //.. ATTTTCGTTTA CGATTTTGTC CGAACCGGAT ACGCCGATTA AGGCGAAGCT
     51  CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTCGGGC GGTTACAACA
    101  GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCGTTT
     151  GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
     201  GGTTGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
     251  TGA AAAATCG CGGCGGCAAG GCGTTTGTGC GCGTGTGGG TGCGGACGGC
     301  AAGGCGGCGG AACCGGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
    151  351  CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGTC ATCTCCGAAA
     401  TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACCGGCCCT AGGCGGCCCG
     451  CCGCGCCGAT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1  MAKMMKWA AVAAAVVWG GWS.LKPEPH VLDITETVRR G.....
      51  .....
     101  .....
     151  .....
     201  ..... I SFTILSEPDT
     251  PIKAKLDSVD PGLTMMSSGG YNSSTDASN AVYYYARSFV PNPDKLATG
    251  301  MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTGM
     351  RDSMNTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*
  
```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1  . .GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA
      51  ACTCGGGCAA CAGGTTAAAA AGGCGGATTT GATTGCGGAA ATCAATTGCA
     101  CCTCGCAGAC CAATACGCTC AATACGGAAA AATCCAAGTT GGAACCGTAT
     151  CAGGCGAAGC TGGTGTGCGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA
     201  ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAAGAGG
     251  ATTTGGAAG CGCGCAGGAT GCGTTTGCCG CCGCCAAAGC CAATGTTGCC
     301  GAGCTGAAGG CTTAATCAG ACAGAGCAA ATTTCCATCA ATACCGCCGA
    351  351  GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG
     401  TGGCGATTCT CGTGAAGAG GGCAGACTG TGAACGCGGC GCAGTCTACG
     451  CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAATGCA
     501  GATTGCCGAG GCGGATATTA CCAAGGTGAA GCGGGGCGAG GATATTTCTG
     551  TTACGATTTT GTCGAACCG GATACGCCGA TTAAGGCGAA GCTCGACAGC
    401  601  GTCGACCCCG GGCTGACCAC GATGTCGTCG GCGGTTACA ACAGCAGTAC
     651  GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTTCG TTTGTGCCGA
     701  ATCCGGACGG CAAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA
     751  ATCGACGGCG TGA AAAATGT GCTGATTATT CCGTCGCTGA CCGTGAAAAA
     801  TCGCGGCGGC AAGCGGTTTG TGCGCGTGT GGGTGCGGAC GGCAAGGCGG
    451  851  CGGAACGCGA AATCCGACC GGTATGAGAG ACAGTATGAA TACCGAAGTA
     901  AAAAGCGGGT TGAAGAGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
     951  CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CTAGGCGGC CCGCCGCGCC
    1001  GATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1  . .VSVGAQASGQ IKILYVKLGQ QVKKGLIAE INSTSQTNTL NTEKSKLETY
      51  QAKLVSAQIA LGSAEKKYKR QAALWKENAT SKEDLESAQD AFAAAKANVA
     101  ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GQTVNAAQST
     151  PTIVQLANLD MMLNKMQIAE GDITKVKAGQ DISFTILSEP DTPIKAKLDS
     201  VDPGLTMMSS GGYNSSTDTA SNAVYYYARS FVENPDGKLA TGMTONTVE
    551  251  IDGVKNVLII PSLTVKNRGG KAFVRVLGAD GKAEREIRT GMRDSMNTEV
     301  KSLKEGDKV VISEITAAEQ QESGERALGG PPRR*
  
```

Computer analysis of this amino acid sequence gave the following results:

-428-

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

```

5      orf85.pep      MAKMMKWA AVAAVAVG GWS-LKPEPHVLDITETVRRG
      orf85a          MAKMMKWA AVAAVAVG GWSYLKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS
      10              20              30              40
10     orf85.pep      .....ISFTILSEPDTPIKAKLDSVDPGLTTMSSG
      orf85a          TIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSSG
      15              210             220             230             240             250             260
15     orf85.pep      GYNSSTD TASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK
      orf85a          GYNSSTD TASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK
      20              270             280             290             300             310             320
20     orf85.pep      AFVRVLGADGKAAEREIRTGMRDSMNTGVKSGLKEGDKVVI SEITAAEQQESGERALGGP
      orf85a          AFVRVLGADGKAAEREIRTGMRDSMNTGVKSGLKEGDKVVI SEITAAEQQESGERALGGP
      25              330             340             350             360             370             380
30     orf85.pep      PRRX
      orf85a          PRRX
      30              390

```

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

```

35      1  ATGGCAAAAA  TGATGAAATG  GGCGGCTGTT  GCGGCGGTCG  CGGCGGCAGC
      51  GGT TTGGGGC  GGATGGTCTT  ATCTGAAGCC  CGAGCCGCAG  GCTGCTTATA
      101  TTACGGAAC  GGTCAGGCGC  GGCGACATCA  GCCGACGGT  TTCTGCAACA
      151  GGGGAGATT  CGCCGTCCAA  CCTGCTATCG  GTCGGCGCGC  AGGCATCGGG
      201  GCAGATTAAG  AAACCTTATG  TCAAACCTCG  GCAACAGGTT  AAAAAGGCGC
      251  ATTTGATTGC  GGAAATCAAT  TCGACCTCGC  AGACCAATAC  GCTCAATACG
      301  GAAAAATCCA  AATTGGAAAC  GTATCAGGCG  AAGCTGGTGT  CGGCACAGAT
      351  TGCATTGGGC  AGCGCGGAGA  AGAAATATAA  GCGTCAGGCG  GCGTTGTGGA
      401  AGGATGATGC  GACCGCTAAA  GAAGATTGG  AAAGCGCACA  GGATGCGCTT
      451  GCCGCCGCCA  AAGCCAATGT  TGCCGAGCTG  AAGGCTCTAA  TCAGACAGAG
      501  CAAAATTCC  ATCAATACCG  CCGAGTCGGA  ATTGGGCTAC  ACGCGCATTA
      45  551  CCGCAACGAT  GGACGGCAGC  GTGGTGGCGA  TTCTCGTGGA  AGAGGGGCGC
      601  ACTGTGAACG  CGGCGCAGTC  TACGCCGACG  ATTGTCCAAT  TGGCGAATCT
      651  GGATATGATG  TTGAACAAAA  TGCAGATTGC  CGAGGGCGAT  ATTACCAAGG
      701  TGAAGGCGGG  GCAGGATATT  TCGTTTACGA  TTTTGTCCGA  ACCGGATACG
      751  CCGATTAAGG  CGAAGCTCGA  CAGCGTCGAC  CCCGGGCTGA  CCACGATGTC
      801  GTCGGGCGGC  TACAACAGCA  GTACGGATAC  GGCTTCCAAT  GCGGTCTACT
      851  ATTATGCCCG  TTCGTTTGTG  CCGAATCCGG  ACGGCAAACT  CGCCACGGGG
      901  ATGACGACGC  AGAATACGGT  TGAAATCGAC  GGTGTGAAAA  ATGTGCTGAT
      951  TATTCGCTCG  CTGACCGTGA  AAAATCGCGG  CGGCAGGGCG  TTTGTGCGCG
      1001  TGTGGGTG  AGACGGCAAG  GCGGCGGAAC  GCGAAATCCG  GACCGGTATG
      55  1051  AGAGACAGTA  TGAATACCGA  AGTAAAAAGC  GGGTTGAAAG  AGGGGGACAA
      1101  AGTGGTCATC  TCCGAAATAA  CCGCCGCCGA  GCAGCAGGAA  AGCGGCGAAC
      1151  GCGCCCTAGG  CGGCCGCGCG  CGCCGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 770>:

```

60      1  MAKMMKWA AVAAVAVG  GWSYLKPEPQ  AAYITETVRR  GDISRTVSAT
      51  GEISPSNLVS  VQAQASGQIK  KLYVKLGQV  KKGDLIAEIN  STSQNTNLNT
      101  ESKKLETYQA  KLVSAQIALG  SAEKKYKRQA  ALWKDDATAK  EDLESAQDAL
      151  AAAKANVAEL  KALIRQSKIS  INTAESELGY  TRITATMDGT  VVAILVEEGQ
      201  TVNAAQSTPT  IVQLANLDM  LNKMQIAEGD  ITKVKAGQDI  SFTILSEPDT
      251  PIKAKLDSVD  PGLTTMSSG  YNSSTDASN  AVYYARFV  PNPDGKLATG
      65  301  MTTQNTVEID  GVKNVLIIPS  LTVKNRGGRA  FVRVLGADGK  AAEREIRTGM

```

351 RDSMNTTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR\*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

5	orf85a.pep	30 40 50 60 70 80	PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE
	orf85-1		VSVGAQASGQIKILYVKLGQQVKKGDLIAE
10	orf85a.pep	90 100 110 120 130 140	INSTSQNTLNTTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATAKEDLESAQD
	orf85-1		INSTSQNTLNTTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD
15	orf85a.pep	150 160 170 180 190 200	ALAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
	orf85-1		AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
20	orf85a.pep	210 220 230 240 250 260	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
	orf85-1		PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
30	orf85a.pep	270 280 290 300 310 320	GGYNSSTDASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
	orf85-1		GGYNSSTDASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
35	orf85a.pep	330 340 350 360 370 380	RAFVRVLGADGKAAEREIRTGMRDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGG
	orf85-1		KAFVRVLGADGKAAEREIRTGMRDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGG
40	orf85a.pep	390	PPRRX
	orf85-1		PPRRX

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

50	ORF85	1	MAKMMKWA AVAAVAVVGGWS.LKPEPHVLDITETVRRG.....	40
	ORF85ng	1	MAKMMKWA AVAAVAVVGGWSYLKPEPQAAYITEAVRRGDISRTVSAT	50
55	ORF85	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDT	250
	ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDT	250
60	ORF85	251	PIKAKLDSVDPGLTTMSSGGYNSSTDASNAVYYYARSFVNPDPGKLATG	300
	ORF85ng	251	PIKAKLDSVDPGLTTMSSGGYNSSTDASNAVYYYARSFVNPDPGKLATG	300
65	ORF85	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM	350
	ORF85ng	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM	350
	ORF85	152	RDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR	393
	ORF85ng	351	KDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR	393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```
1 ATGGCAAAAA TGATGAAATG GCGGGCTGTT GCGGCGGTCG CGGCGGCaac
51 GGTTTGGGGC GGATGGTCTT ATCTGAAGCC CGAACCGCAG GCTGCTTATA
101 TTACGGAaac ggTCAGGCGC GCGGATATCA GCCGGACGGT TTCCGCGACG
151 GgcgAGATTT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
201 GCAGATTAAA AAGCTTTATG TCAAACTCGG GCAACAGGTC AAAAAGGGCG
251 ATTTGATTGC GGAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301 GAAAAATCCA AATTGGAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCCTCAGGCG GCGTTGTGGA
401 AGGATGATGC GACCTCTAAA GAAGATTGGA AAAGCGCGCA GGATGCGCTT
451 GCCCGCGCCA AAGCCATGT TGCCGAGTTG AAGGCTTTAA TCAGACAGAG
501 CAAAATTTCC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATT
551 CCGCGACGAT GGACGGGACG GTGGTGGCGA TTCCCGTGGA AGAGGGGCAG
601 ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
15 651 GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
701 TGAAGCGGGC GCAGGATATT TCGTTTACGA TTTTGTCCGA ACCGGATACG
751 CCGATTAAAG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTATT
851 ATTATGCCCG TTCGTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
20 901 ATGACGACGC AGAATACGGT TGAAATCGAC GGTGTGAAAA ATGTGTGCT
951 TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAAGGCG TTCGTACGCG
1001 TGTTGGGTGC GGACGGCAAG GCAGTGGAA CCGAAATCCG GACCGGTATG
1051 AAAGACAGTA TGAATACCGA AGTGAAAAGC GGGTTGAAAG AGGGGGACAA
1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
25 1151 GCGCCCTAGG CGGCCCGCCG CGCCGATAA
```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```
1 MAKMMKWA AVAAAVWG GWSYLKPEPQ AAYITEAVRR GDISRTVSAT
51 GEISPSNLVS VGAQASGQIK KLYVKLGQOV KKGDLIAEIN STTQNTNIDM
101 ESKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATSK EDLESAQDAL
30 151 AAAKANVAEL KALIRQSKIS INTAESDLGY TRITATMDGT VVAIPVEEGQ
201 TVNAAQSTPT IVQLANLDM LNMQIAEGD ITKVKGQDI SFTILSEPDT
251 PIKAKLDSVD PGLTTMSSG YNSSTDASN AVYYYARFV PNPDGKLATG
301 MTTQNTVEID GVKNVLLIPS LTVKNRGGKA FVRVLGADGK AVEREIRTGM
351 KDSMNTEVKS GLKEGDKVVI SEITAEQQE SGERALGGPP RR*
```

ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```
30 40 50 60 70 80
orf85ng PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQOVKKGDLIAE
40 orf85-1 VSVGAQASGQIKILYVKLGQOVKKGDLIAE
10 20 30

90 100 110 120 130 140
orf85ng INSTTQNTNIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEDLESAQD
45 orf85-1 INSTSQNTNINTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD
40 50 60 70 80 90

150 160 170 180 190 200
orf85ng ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST
50 orf85-1 AFAAAKANVAELKALIRQSKISINTAESELYTRITATMDGTVVAILVEEGQTVNAAQST
100 110 120 130 140 150

210 220 230 240 250 260
orf85ng PTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDTPIKAKLDSVDPLTTMSS
55 orf85-1 PTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDTPIKAKLDSVDPLTTMSS
160 170 180 190 200 210

270 280 290 300 310 320
orf85ng GGYNSSTDASNNAVYYYARFVNPDPGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
60 orf85-1 GGYNSSTDASNNAVYYYARFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
220 230 240 250 260 270

330 340 350 360 370 380
```

-431-

```

orf85ng      KAFVRVLGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVISSEITAAEQQESGERALGG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf85-1      KAFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVISSEITAAEQQESGERALGG
                280      290      300      310      320      330

orf85ng      390
              PPRRX
              |||||
orf85-1      PPRRX

```

10 In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from  
membrane fusion protein precursor, MTRC\_NEIGO SW: P43505 (412 aa) [Escherichia  
coli] Length = 380

Score = 193 bits (485), Expect = 2e-48

Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

```

Query: 29  PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDIAE 88
          P   Y T  VR GD+ ++V ATG++      V VGAQ SQQ+K L V +G +VKK  L+
Sbjct: 41  PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTLSSVAIGDKVKKQDLGV 100

```

```

Query: 89  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKYKROAALWKDDATSKEXXXXXXX 148
          I+   N I  ++ L  +A+  A+  L  A  Y RQ  L +  A S++
Sbjct: 101 IDFEQAENQIKEVEATLMELRAQRQQAELKLARVTYSRQORLAQTKAVSQDLDLTAAT 160

```

```

Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVVAIPVEEGOTVNAAOST 208
          I++++ S++TA+++L YTRI A M G V I  +GQTV AAQ
Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLDATAKTNLDYTRIVAPMAGEVTQITTLQGGQTVIAAQQA 220

```

```

Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTMTSS 268
          P I+ LA++ ML K Q++E D+  +K GQ  FT+L +P T  + ++ V P
Sbjct: 221 PNILTLADMSAMLVKAQVSEADVIHLKPGQKAWFTVLGDELTRYEGQIKDVLP----- 273

```

```

Query: 269 GGYNSSTDASNAVYYIARFVNPDPGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328
          + +  ++A++YYAR VNP+G L  MT Q  +++ VKNVL IP  + +  G
Sbjct: 274 -----TPEKVNDAIFYIARFEVNPENGLRLDMDTAQVHIQLTDVKNVLTIPLSALGDFVG 328

```

```

Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVISSE 372
          +V L  +G+  ERE+  G ++  + E+  GL+ GD+VVI E
Sbjct: 329 DNRYKVKLLRNGETREREVTIGARNDTVEIVKGLEAGDEVVIGE 373

```

40 Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*,  
and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A  
shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein

45 was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis  
(Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a  
surface-exposed protein, and that it is a useful immunogen.

### Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

50      1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
      51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
     101  CGGTTGTCGG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
     151  GGCAAACTGT ATGCGGAAGc CAAATTCGCC GACGcAGCG TAACTTACGG
     201  CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAAGGCT ATGGATTGTG

```

-432-

251 TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCCGGGG  
 301 CTGAAAATCA CCAACGGCAA AAACTTTTAT TCCGTCGGCG GTTTGAATAA  
 351 GGCGGGTACA GGAATAATACA GCATAGGCGG CGTGGAAACC GAAGTCGTCA  
 401 AATATCGGGT GCGGCGCGGC GACGATGCGG TAATGTATTT cTTGCACCG  
 5 451 TCCCTGAACA ATATTCGGGC ACAAATCGGC TATACCGACG ACGGCAAAAC  
 501 CTATACGCTG AAACCTCAAAT CGGTGCAGAT CAACGGCCAG GCAGCCAAAC  
 551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYRDIRR  
 10 51 GKLYAEAKFA DGSVTYKGAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG  
 101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP  
 151 SLNNIPAQIG YTDDGKTYTL KLSVQINGQ AAKP\*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

1 ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC  
 15 51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT  
 101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC  
 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG  
 201 TTTCGAGTCC GGCGGTACGG TTGTCGGCAA TACCCTGCAC CCTACCTACT  
 251 ATAGAGACAT ACGCAGGGGC AAACCTGTATG CGGAAGCCAA ATTCGCCGAC  
 20 301 GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAGCCCC  
 351 CAAGGCTATG GATTGTGTTCA CGCTTGCTTG GCAGTTGGCG GCAAAATGACG  
 401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC  
 451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT  
 501 GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA  
 25 551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT  
 601 ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA  
 651 CCGCCAGGCA GCCAAACCGT AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

1 MMKTFKNIFS AAILSAAALPC AYAAGLPQSA VLHYSYSGYI PATMTFERSG  
 30 51 NAYKIVSTIK VPLYNIRFES GGTVVGNLTH PTYYRDIRRG KLYAEAKFAD  
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS  
 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY  
 201 TDDGKTYTLK LKSVQINGQA AKP\*

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N. meningitidis*:

					10	20	30
40	orf120.pep				IPATMTFERSGNAYKIVSTIKVPLYNIRFE		
	orf120a	SAAILSAAALPCAYAAGLPXS	AVLHYSYSGYIPAT	XXXXXXXXNAXKIVSTIKVPLYNIRFE			
		10	20	30	40	50	60
45	orf120.pep		40	50	60	70	80
	orf120a	SGGT	VVGNTLHPTYYRDIRRGKLYAEAKFADG	SVTYGKAGESKTEQSPKAMD	DLFTLAWQL		
		70	80	90	100	110	120
50	orf120.pep		100	110	120	130	140
	orf120a	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE	EVVKYRVRRGDDAVMYFFAP				
55		130	140	150	160	170	180

-433-

```

              160      170      180
orf120.pep  SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
            |||
orf120a     SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
5           190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

1  ATGATGAAGA CTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT
101 ATTCGGGCAG CTACGGCATT CCCGCCACNA NNANNTNNGN ACNNNGNGNC
151 AATGCTTNCA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201 TTTCGAGTCC GCGGTACGG TTGTCCGCAA TACCCTGCAC CCTACCTACT
251 ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC
301 GGCAGCGTAA CCTACGGCAA AGCGNNNNN ANCNNNNNG NGCAAAGCCC
351 CAAGGCTATG GATTGTGTTCA CGCTGCGNTG GCAGTTGGCG GCAAATGACG
15 401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAA ACTTTATTCC
451 GTCGGCGGTT TGAATAAGGC GGTACAGGA AAATACAGCA TAGGCGGCGT
501 GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA
551 GTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601 ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
20 651 CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

1  MMKTFKNIFS AAILSAALPC AYAAGLPXSA VLHYSGSYGI PATXXXXXXX
51 NAXKIVSTIK VPLYNIRFES GGTVVGNLTH PTYYRDIRRG KLYAEAKFAD
101 GSVTYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
25 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAIQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

              10      20      30      40      50      60
30 orf120a.pep  MMKTFKNIFS AAILSAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIK
orf120-1       MMKTFKNIFS AAILSAALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
              10      20      30      40      50      60

              70      80      90      100     110     120
35 orf120a.pep  VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYKAXXXXXXQSPKAM
orf120-1       VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYKAGESKTEQSPKAM
              70      80      90      100     110     120

              130     140     150     160     170     180
40 orf120a.pep  DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
orf120-1       DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
              130     140     150     160     170     180

              190     200     210     220
45 orf120a.pep  DAVMYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
orf120-1       DAVMYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
50           190     200     210     220

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55 orf120.pep                                IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
orf120ng    SAAILSAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIKVPLYNIRFE 69

60 orf120.pep  SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYKAGESKTEQSPKAMD LFTLAWQL 90
orf120ng     SGGTVVGNLHPAYYKDIRRGKLYAEAKFADGSVTYKAGESKTEQSPKAMD LFTLAWQL 129

```

-434-

5 orf120.pep AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP 150  
 orf120ng AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDTVTYFFAP 189  
 orf120.pep SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 184  
 orf120ng SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 223

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

10 1 ATGATGAAGA CTTTTTAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC  
 51 CCTGCCGTGC GCGTATGCGG CAAGGCTACC CCAATCCGCC GTGCTGCACT  
 101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC  
 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG  
 201 TTTTCAATCC GCGGTACGG TTGTCGGCAA TACCCTGCAC CTGCCTACT  
 15 251 ATAAAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCCGCCGAC  
 301 GGCAGCGTAA CCTACGGCAA AGCGGGCGAG AGCAAACCG AGCAAAGCCC  
 351 CAAGGCTATG GATTTGTTCA CGCTTGCCGT GCAGTTGGCG GCAAATGACG  
 401 CGAAACTCCC CCCGGTCTG AAAATCACCA ACGGCAAAA ACTTTATTCC  
 451 GTCGGCGGCC TGAATAAGGC GGTACGGGA AAATACAGCA TaggCGGCGT  
 20 501 GGAAACCGAA GTCGTCAAAT ATCGGGTGC GCGCGCGCAC GATACGGTAA  
 551 CGTATTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT  
 601 ACCGACGACG GCAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA  
 651 CGGACAGGCC GCCAAACCGT AA

This encodes a protein having amino acid sequence <SEQ ID 780>:

25 1 MMKTFKNIFS AAILSAALPC AYAARLPQSA VLHYSYSGYGI PATMTFERSG  
 51 NAYKIVSTIK VPLYNIRFES GGTVVGNLTH PAYYKDIRRG KLYAEAKFAD  
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS  
 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAQIGY  
 201 TDDGKTYTLK LKSVQINGQA AKP\*

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

35 orf120-1.pep 10 20 30 40 50 60  
 MMKTFKNIFS AAILSAALPC AYAAGLPQSA VLHYSYSGYGI PATMTFERSG NAYKIVSTIK  
 orf120ng MMKTFKNIFS AAILSAALPC AYAARLPQSA VLHYSYSGYGI PATMTFERSG NAYKIVSTIK  
 40 orf120-1.pep 70 80 90 100 110 120  
 VPLYNIRFES GGTVVGNLHPTYYRDIRRGKLYAEAKFAD GSVTYGKAGESKTEQSPKAM  
 orf120ng VPLYNIRFES GGTVVGNLHPAYYKDIRRGKLYAEAKFAD GSVTYGKAGESKTEQSPKAM  
 45 orf120-1.pep 130 140 150 160 170 180  
 DLFTLAWQLA ANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD  
 orf120ng DLFTLAWQLA ANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD  
 50 orf120-1.pep 190 200 210 220  
 DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX  
 orf120ng DTVTYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX

This analysis, including the presence of a putative leader sequence in the gonococcal protein  
 55 suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be  
 useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:



-435-

1 ATGTATCGGA GGAAAGGGCG GGCATCAAG CCGTGGATGG GTGCCGGTGC  
 51 .GCGTTTGCC GCCTTGGTCT GGCTGGTTT CGCGCTCGGC GATACTTTGA  
 101 CTCGGTTTGC GGTGCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT  
 5 201 GATGGTGTTC TCCTTGATT TGTGTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGGT CGGCAGTTC AACAAATTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTCAGGCGC  
 10 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATT..

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL  
 151 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNI..

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

1 ATGTATCGGA GGAAAGGGCG GGCATCAAG CCGTGGATGG GTGCCGGTGC  
 51 GCGTTTGCC GCCTTGGTCT GGCTGGTTT CGCGCTCGGC GATACTTTGA  
 101 CTCGGTTTGC GGTGCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT  
 20 201 GATGGTGTTC TCCTTGATT TGTGTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGGT CGGCAGTTC AACAAATTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 25 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTCAGGCGC  
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGTGCTTCC  
 501 CTTGCTGCTT TACTATTTCC TGCTGGATTG GCAGCGGTGG TCGTGGCGCA  
 551 TTGCCAAACT GGTCCGAGG CGTTTTCGCG GTGCTTATAC GCGCATTACA  
 601 GGCAATTTGA ACGAGGTATT GGGCGAATT TCGCGCGGGC AGCTTCTGGT  
 30 651 AATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGTTG CTGGTCGGGC  
 701 TGGATTCGGG GTTTGCCATC GGTATGCTTG CCGGTATTTT GGTGTTTGTG  
 751 CCTTATCTCG GGGCGTTTAC GGGATTGCTG CTGCCACCG TCGCCGCCTT  
 801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTTTG  
 851 CCGTAGGACA GTTCTCGAA AGTTTTTCA TTACGCCGAA AATCGTGGGA  
 35 901 GACCGTATCG GGCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT  
 951 CGGGCAGCTG ATGGGCTTGG TCGGAATGTT GCGGGGATTG CCTTTGGCCG  
 1001 CCGTAACCTT GGTCTTGCTT CCGGAGGGCG TGCAGAAATA TTTTGCCGGC  
 1051 AGTTTTTACC GGGCAGGTA G

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL  
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNIVSSI GNLLLLPLLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT  
 201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAT GMLAGILVEV  
 45 251 PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG  
 301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAAVTLVLI REGVQKYFAG  
 351 SFYRGR\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
orf121.pep	MYRRKGRGIK	PWMGAGXAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR	
55 orf121a	MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR	
		10	20	30	40	50	60
		70	80	90	100	110	120
orf121.pep	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL	IGFMQNTLLP	WLKNTIGGYV	

-436-

```

|||||
orf121a  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
          70      80      90      100     110     120

5      130      140      150
orf121.pep EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI
          |||||
orf121a  EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
          130      140      150      160      170      180

10     190      200      210      220      230      240
orf121a  SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI

```

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

```

15      1  ATGTATCGGA  GGAAAGGGCG  GGCATCAAG  CCGTGGATGG  ATGCCGGTGC
51      51  GCGGTTTGCC  GCCTTGGTCT  GGCTGGTTT  CGCGCTCGGC  GATACTTTGA
101     101  CTCCGTTTGC  GGTTGCGGCG  GTGCTGGCGT  ATGTATTGGA  CCCTTTGGTC
151     151  GAATGGTTGC  AGAAAAAGGG  TTTGAACCGT  GCATCCGCTT  CGATGCTGTG
201     201  GATGGTGTTT  TCCTTGATTT  TGTGTGGC  ATTATTGTG  ATTATTGTCC
251     251  CTATGCTGGT  CGGGCAGTTC  AACAAATTGG  CATCGCGCCT  GCCCCAATTA
20      301  ATCGGTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAA  ATACAATCGG
351     351  CGGATATGTG  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  CTTCAGGCGC
401     401  ATACGGGCGA  GTTGAGCAAC  GCGCTTAAGG  CGTGGTTTCC  CGTTTTGATG
451     451  AGGCAGGCG  GCAATATTGT  CAGCAGTATC  GGCAACCTGC  TGCTGCTTCC
501     501  CTTGCTGCTT  TACTATTTC  TGCTGGATTG  GCAGCGGTGG  TCGTGCGGCA
25      551  TTGCCAACT  GGTTCGAGG  CGTTTTGCCG  GTGCTTATAC  GCGCATTACA
601     601  GGCAATTGA  ACGAGGTATT  GGGCGAATTT  TTGCGCGGGC  AGCTTCTGGT
651     651  GATGCTGATT  ATGGGTTTGG  TTTACGGCTT  GGGGTTGGTG  CTGGTCGGGC
701     701  TGGATTCGGG  GTTTGCAATC  GGTATGGTTG  CCGGTATTTT  GGTTTTTGTT
751     751  CCCTATTTGG  GCGCGTTTAC  AGGACTGCTG  CTGGCAACCG  TCGCCGCCTT
30      801  GCTCCAGTTC  GGTTCTGGA  ACGGCATCTT  GGCTGTTTGG  GCGGTTTTTG
851     851  CCGTAGGACA  GTTCTCGAA  AGTTTTTCA  TTACGCCGAA  AATCGTGGGA
901     901  GACCGTATCG  GCCTGTCGCC  GTTTTGGGTT  ATCTTTTCGC  TGATGGCGTT
951     951  CGGGCAGCTG  ATGGGCTTTG  TCGGAATGTT  GGCCGGATTG  CTTTGGCCG
1001    1001  CCGTAACCTT  GGTCTTGCTT  CGCGAGGCG  TGCAGAAATA  TTTTGCCGGC
35      1051  AGTTTTTACC  GGGGCAGGTA  G

```

This encodes a protein having amino acid sequence <SEQ ID 786>:

```

40      1  MYRRKGRGIK  PWMDAGAAFA  ALVWLVFALG  DTLTPFAVA  VLAYVLDPLV
51      51  EWLQKKGLNR  ASASMSVMVF  SLILLALLL  IIVPMLVGQF  NNLASRLPQL
101     101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  LQAHTGELSN  ALKAWFPVLM
151     151  RQGGNIVSSI  GNLLLLPLLL  YYFLLDWQRW  SCGIAKLVER  RFAGAYTRIT
201     201  GNLNEVLGEF  LRGQLLVMLI  MGLVYGLGLV  LVGLDSGFAI  GMVAGILVFV
251     251  PYLGAFTGLL  LATVAALLQF  GSWNGILAVW  AVEAVGQFLE  SFITPKIVG
301     301  DRIGLSPFWV  IFSLMAFGQL  MGFVGMLAGL  PLAAVTLVLL  REGVQKYFAG
351     351  SFYRGR*

```

45 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

```

50      10      20      30      40      50      60
orf121a.pep  MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
          |||||
orf121-1     MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
          10      20      30      40      50      60

          70      80      90      100     110     120
orf121a.pep  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
          |||||
55      70      80      90      100     110     120
orf121-1     ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
          70      80      90      100     110     120

          130     140     150     160     170     180
orf121a.pep  EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
          |||||
60      130     140     150     160     170     180
orf121-1     EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
          130     140     150     160     170     180

          190     200     210     220     230     240
65      190     200     210     220     230     240
orf121a.pep  SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI

```

-437-

	orfl21-1	SCGI AKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI	190	200	210	220	230	240
5	orfl21a.pep	GMVAGILVFVYPYLGAFTGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESEFFITPKIVG	250	260	270	280	290	300
	orfl21-1	GMLAGILVFVYPYLGAFTGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESEFFITPKIVG	250	260	270	280	290	300
10	orfl21a.pep	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX	310	320	330	340	350	
15	orfl21-1	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX	310	320	330	340	350	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from *N.gonorrhoeae*:

20	orfl21.pep	MYRRKGRGIKPWMGAGXAFALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR	60
	orfl21ng	MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTPFAVAAYVLDPLVEWLQKKGLNR	60
25	orfl21.pep	ASASMSVMVFSLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
	orfl21ng	ASASMSVMVFSLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
	orfl21.pep	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI	156
30	orfl21ng	EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSTIGNLLLPPLLLYYFLLDWHRW	180

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEQ ID 788>:

	1	MYRRKGRGIK	PWMGAGAAFA	ALVWLVYALG	DTLTPFAVAA	VLAYVLDPLV
35	51	EWLQKKGLNR	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL
	101	IGFMQNTLLP	WLKNTIGGYV	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM
	151	KQGGNIVSTI	GNLLLPPLLL	YYFLLDWHRW	SCGIPKLVR	RFAGAYTRIT
	201	GNLNKVGWKF	LRGQLLGETE	RGAVVCRVGR	ECWEGGGARS	RPSDDGWPRW
	251	GGG*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 789>:

40	1	ATGTATCGGA	GAAAAGGACG	GGGCATCAAG	CCGTGGATGG	GTGCCGCGCG
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTA	CGCGCTCGGC	GATACTTTGA
	101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTGTTGGA	CCCTTTGGTC
	151	GAATGTTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGTCTGT
45	201	GATGGTGTTC	TCCTTGATT	TGTTGTTGGC	ATTATTGTTG	ATTATTGTCC
	251	CTATGCTGGT	CGGGCAGTTC	AATAATTTGG	CATCTCGCCT	GCCCCAATTA
	301	ATCGGTTTGA	TGCAGAACAC	GCTGCTGCCG	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	TTTCAGGCGC
	401	ATACGGGCGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTGTATG
	451	AAACAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCCGCC
50	501	CTTGCTGCTT	TACTATTTC	TGCTGGATTG	GCAGCGGTGG	TCGTGCGGCA
	551	TCGCCAAACT	GGTTCCGAGG	CGTTTGGCCG	GTGCTTATAC	GCGCATTACG
	601	GGTAATTTGA	ACGAGGTATT	GGGCGAATTT	TTGCGCGGTC	AGCTTCTGGT
	651	GATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTGATG	CTAGTCGGAC
	701	TGGATTCGGG	ATTTGCCATC	GGTATGGTTG	CCGGTATTTT	GGTGTGTGTC
55	751	CCCTATTTGG	GTGCGTTTAC	GGGATTGCTG	CTTGCCACTG	TTGCAGCCTT
	801	GCTCCAGTTC	GGTTCGTGGA	ACGGAATCTT	GGCTGTTTGG	GCGGTTTTTG
	851	CCGTCGGTCA	GTTTCTCGAA	AGTTTTTTCA	TTACGCCGAA	AATTGTAGGA
	901	GACCGTATCG	GCCTGTCGCC	GTTTTGGGTT	ATCTTTTCGC	TGATGGCGTT
	951	CGGAGAGCTG	ATGGGCTTTG	TCGGAATGTT	GGCCGGATTG	CCTTTGGCCG
60	1001	CCGTAACCTT	GGTCTTGCTT	CGCAGGGCGC	CGCAGAAATA	TTTTGCCCGG
	1051	AGTTTTTACC	GGGGCAGGTA	G		

-438-

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

```

1 MYRRKGRGIK PWMGAGAAFA ALVWLVAALG DTLTPFAVAA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
151 KQGGNIVSSI GNLLPPLL YYFLLDWQRW SCGIKLVPR RFAGAYTRIT
201 GNLNEVLGEF LRGLLVMLI MGLVYGLGLM LVGLDSGFAI GMVAGILVVFV
251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
301 DRIGLSPFWV IFSLMAFGEL MGFVGMLAGL PLAAVTLVLL REGAQKYFAG
351 SFYRGR*

```

ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:

```

10 20 30 40 50 60
orf121-1.pep MYRRKGRGIKPWMGAGAAFAALVWLVAALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
15 orf121ng-1 MYRRKGRGIKPWMGAGAAFAALVWLVAALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
10 20 30 40 50 60

70 80 90 100 110 120
orf121-1.pep ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
20 orf121ng-1 ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
70 80 90 100 110 120

130 140 150 160 170 180
orf121-1.pep EIDQASIIAWLQHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLPLLLYFLLDWQRW
25 orf121ng-1 EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSSIGNLLLPLLLYFLLDWQRW
130 140 150 160 170 180

190 200 210 220 230 240
orf121-1.pep SCGIKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI
30 orf121ng-1 SCGIKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI
190 200 210 220 230 240

250 260 270 280 290 300
orf121-1.pep GMLAGILVFPYLPYLGAFGLLATVAALLQFSGWNGILSVWAVFAVGQFLESFFITPKIVG
35 orf121ng-1 GMVAGILVFPYLPYLGAFGLLATVAALLQFSGWNGILAVWAVFAVGQFLESFFITPKIVG
250 260 270 280 290 300

310 320 330 340 350
orf121-1.pep DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
40 orf121ng-1 DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX
310 320 330 340 350

```

In addition, ORF121ng-1 shows homology to a permease from *H.influenzae*:

```

sp|P43969|PERM HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
Score = 69.9 bits (168), Expect = 2e-11
Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)
50 Query: 26 VYALGDTLTPFAVAAVLAYVLDPLVEWL-QKKGLNRASASMSVMVFXXXXXXXXXXV 84
+Y GD + P +A VL+Y+L+ + +L Q R A++ + VP
Sbjct: 32 IYFFGDLIAPLLIALVLSYLLEIPINFLNQYLCPRMLATILIFGSFIGLAAVFFLVLP 91

55 Query: 85 MLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNALK 143
ML Q +L S LP + N WL N Y E ID + + + F + ++ +
Sbjct: 92 MLWNQTISLLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFGE 147
Query: 144 AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXDWQRWSCGIKLVPRRFAGAYTRITGNL 203

60 + + + N+VS D G+++ +P+ A+ R +
Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRFLPKNRNLAFXRWK-EM 206

Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXDSGFAIGMVAGILVFPYXXXXXXXXXXXX 263
+ + ++ G+ + + G+ V VPY
65 Sbjct: 207 QQQISNYIHGKLEILIVTLITYIIFLIFGLNYPLLLAFVGLSVLPYIGAVIVTIPVA 266

```

Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323  
 QFG + FAV Q L+ + P + + + L P +I S++ FG L GF  
 Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIIIISVLIFGGLWGF 326

5 Query: 324 VGMLAGLPLAAVTLVLL 340  
 G+ +PLA + ++  
 Sbjct: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and  
 10 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

15 1 ..ACTGCTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT  
 51 TTTGTCCTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT  
 101 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
 151 CTGCGCCTCT ATGCCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG  
 201 TTTGCTTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
 20 251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTTCGTGTC  
 301 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC  
 351 TTTGATACGC CGCACGCAA AGGACGCGC TGTCCGCATC TTTGAACTCT  
 401 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
 501 CGAGCAGCCC TTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAG..

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

1 ..TAFSAALRLS PSXLVIFLSF GKPYYQTAAI LTFFCTSCPP RSNAYQQYRR  
 51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRRECGFLC  
 101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCT  
 151 EQRVGNVQQ RIGIGVSEQP FFKWDFNSAK YQ..

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

1 ATATCGTACT GGGCAAGCAG TTCGCCGAT TTTTGGAG TAGATACCGC  
 51 GCCTTTGATT TTTTGCCGC TCTACCCAA GGCTTCGATG AAAAAGTTGA  
 101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTTCGGG TACGAATTCG  
 35 151 ACTGCTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTGTG TCGTCATATT  
 201 TTTGTCCTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT  
 251 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
 301 CTGCGCCTCT ATGCCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG  
 351 TTTGCTTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
 40 401 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTTTGG GTTTCGTGTC  
 451 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC  
 501 TTTGATACGC CGCACGCAA AGGACGCGC TGTCCGCATC TTTGAACTCT  
 551 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 601 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
 651 CGAGCAGCCC TTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAGCTTT  
 45 701 CTGCCTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT  
 751 CGTCATCGTT TGTGTTCTG A

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

1 ISYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMEPVPM PIYSFSGTNS  
 50 51 TAFSAAMRLS SSCVVIFLSF GKPYYQTAAI LTFFCTSCPP RSNAYQQYRR  
 101 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRREFGFLC  
 151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCT  
 201 EQRVGNVQQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDDTV  
 251 RHRLCS\*

Computer analysis of this amino acid sequence gave the following results:

[illegible]

	1	ATATCATATT	GGGCAAGCAG	TTCAC'TGGAT	TTTTTGGAA	TAGATACCGC
30	51	GCCTTTGATT	TTTTTGGCGC	TCTTACCCAA	GGCTTCGATG	AAAAAGTTGA
	101	TGGTCGAAC	GGTACCGATG	CCGATGTATT	CGTTTTTCGG	TACGAATTCG
	151	ACTGCNTTTT	CGGCGGCGAT	CGCGTTGAGT	TCGCTCTGTG	TCGTCATATT
	201	TTTGTCTCTT	GGGAAACCGT	ATCAACAAC	AGCCGCCATC	TTAACATTTT
	251	TTNNNACGTC	CTGCCGCGCG	CGTTCAAATC	CTTACCAGCA	ATACCGCCGC
35	301	CTGCGACTCT	ATGCCTTCCA	TGCGCCCGAG	ATAACCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GANGTTTGAGC	CACGAAATGT	CTATGCCCAA	ATCGGCCGAC
	401	ATGTTGGCAC	GCA'TTTGCGG	AATATGCGGC	CGAGATTGG	GTTTCTGTGC
	451	AATCAGCGTC	GTATCGACAT	TGACCGCCTG	CCAACCCCTG	GCCTGAACCT
	501	TTTGATACGC	CGCACGCAA	AGGACGCGGC	TGTCGCCATC	TTTGAAC'TC
40	551	CGCGCGGTGT	CGGGGAAATG	GCTGCCGATA	TCGCCCAAAC	CTGCCGCCAC
	601	GAGCAGCGCG	TCGGTAACGG	CGTGCAGCAG	CGCATCGGCA	TCGGAGTGTC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAGTTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATGTT
	751	CGTCATCGTT	TGTGTTCTCTG	A		

45	1	ISYWASSSLD	FLEVDTAPLI	FLPLLPKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVFLSF	GKPYQQTAAI	LTFXFTSCPP	RSNPYQGYRR
	101	LRLYAFHAFE	ITFEVGFVAF	XVDARNVYQA	IGDVGTHLR	NMRREFGQLR
	151	NHGRIDIDRL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGEM	AADIAQTCRT
	201	EQRVNGVQVQ	RIGIGVSEQP	FFKWDFNSAK	YQLSAFGQLV	DIVALSDTDV
50	251	RHRLCS*				

		10	20	30	40	50	60
	orf122a.pep	ISYWASSSLDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPMYSFSGTNSTAFSAAMRLS					
55	orf122-1	ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPPIYSFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
		70	80	90	100	110	120
60	orf122a.pep	SSCVVIFLSFGKPYQQTAAILTFEFTSCPPRSNPYQQYRRLRLYAFHAPETIEFFVGFAF					
	orf122-1	SSCVVIFLSFGKPYQQTAAILTFEFTSCPPRSNAYQQYRRLRLYAFHAPPEIAEFFVGFAF					

-441-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf122a.pep	XVDARNVYAQIGGDVGTHLRNMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122-1	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
10	orf122a.pep	190	200	210	220	230	240
	orf122-1	FELCGGVGEMAADIAQTCRTEQRVGNGVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
15	orf122a.pep	250					
	orf122-1	DIVALSDTDVRHRLCSX					
20		250					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

25	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLPKASMKKLMVEPVPMYSFSGTNSTAFSAAMRLSSSCVIFLSFGKPYQQTAAI	80
	orf122.pep	LTFECTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
30	orf122ng	LTFECTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
35	orf122ng	NVRCEFGFLCNHGRIDIDHLPDLRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
	orf122.pep	EQRVGNGVQQRIGIGVSEQPFFKWFDFNSAKYQ	182
	orf122ng	EQRVGNGVQQRVGIRMPEQPPFFKWFDFNSAKYQLSAFGQLVDIVALSDTDIRHRLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TTCCGCCGAT	TTTTTGGAGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTGTCCCAA	GGCTTCGATG	AAGAAATTGA
	101	tgTTCGAACC	GgtaCCGATG	CCGATGTATT	CGTTTTCCGG	TACGAATTTCG
	151	ACTGCTTTTT	CGGCGGCGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
	201	TTTAtccttt	gGGAaacct	atcaAcaAAc	agcgcgccatC	TTAACATTTT
45	251	TTTGACGtc	ctggccgcgcg	cgttcaAATc	cgtaccaGca	ataccgcgcg
	301	ctgcgcctCT	AtgcCTTCCA	TCCGCCCGAG	ATAGCCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GATatTGACG	CACGAAATAT	CGatacCCaA	atcggcgcGCG
	401	ATGTTGGCAC	GCATTTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTCCTGTGC
	451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
50	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAACTCT
	551	GCGGCGGTGT	CGGAAAATG	GCTGCCGATG	TCGCCCAAAC	CTGCCGCACC
	601	GAGCAGCgcg	tcggtaaCGG	CGTGACGAG	cgcgTcgGCA	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAAT	GGATTTCAA	CTCGCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
55	751	CGTCATCGTT	TGTGTTCTGT	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLLPKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFECTSWPP	RSNPYQQYRR
60	101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDHL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT
	201	EQRVGNGVQQ	RVGIRMPEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSDDI
	251	RHRLCS*				

-442-

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVDTAFLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS					
5	orf122ng	MSYRASSSPDFLEVETAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
	orf122-1.pep	SSCVVIFLSFGKPYQQTAAILTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF					
10	orf122ng	SSCVVIFLSFGKPYQQTAAILTFFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAF					
		70	80	90	100	110	120
	orf122-1.pep	DVDARNVYAQIGGDVGTHLRNVREFFGLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
15	orf122ng	DIDARNIDTQIGGDVGTHLRNVREFFGLCNHGRIDIDHLPPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
	orf122-1.pep	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
20	orf122ng	FELCGGVGKMAADVAQTCRTEQRVGNVQQRVIRMPEQFFKWFDFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
	orf122-1.pep	DIVALSDTDVRHRLCSX					
25	orf122ng	DIVALSDTDIRHRLCSX					
		250					
30							

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGGCGCGA	GTGCGAACAA	CATTTCGCG	CGTTTTCGCG	AAACACCCGT
	51	CGCTGTCAGC	GTTACCGTGA	TGCGCACGGT	ACTTGCCGTC	ATGCTGCCCG
	101	TTACCGAATA	TGAAACTTC	CTGCTGCTTA	TGCGCTCGGT	ATTGCGCCG
	151	ATGGGGCGGA	TTTGATTGC	CGACTTTTC	GTCTTGAAAC	GGCGTGA

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

40	1	..AGASANNISA	RFAETPVAVS	VTLIGTVLAV	MLPVTEYENF	LLIGSVFAP
	51	MGGFDCRLFR	LETA*			

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTCGGGCA	ATGCCTCCTC	TCCTTCATCT	TCCTCCGCCA	TCGGGCTGAT
	51	TTGGTTCCGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCAGC	GGTACGCTGC
45	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CGGCTCTACT	TTTGGGTCAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTGCGGCG	GCGTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TGATGGGAAA	GCGTGCGCCT	GTCGTTCCGC	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATTT	ACGCCGCGCG	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
50	351	GTGGGACGGC	GAATCTTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGTGA
	401	TTGTGCTGTG	GCTGGTTTTT	GGCGCAGCA	AAACAGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GTTGGCGGTT	CTGTGGCTGA	GTGCCGAAGT
	501	CTTTTCCACG	GCAGGCAGCA	CCGCCGCACA	GGTTTCAGAC	GGCATGAGTT
	551	TCGGAACGGC	AGTCGAGCTG	TCCGCCGTGA	TGCCGCTTTC	CTGGCTGCCG
55	601	CTTGCCGCCG	ACTACACGCG	CCACGCGCGC	CGCCGTTTGG	CGGCAACCCT
	651	GACGGCAACG	CTCGCTTACA	CGCTGACCGG	CTGCTGGATG	TATGCCTTGG
	701	GTTTGGCAGC	GGCGTTGTTT	ACCGGAGAAA	CCGACGTGGC	AAAAATCCTG
	751	CTGGGCGCAG	GTTTGGGTGC	GGCAGGCATT	TTGGCGGTGC	TCCTCTCCAC



-443-

5  
 801 CGTTACCACA ACGTTTCTCG ATGCCTATTC CGCCGGCGCG AGTGCGAACA  
 851 ACATTTCGCG GCGTTTTGCG GAAACACCCG TCGCTGTCGG CGTTACCCTG  
 901 ATCGGCACGG TACTTGCCGT CATGCTGCCC GTTACCGAAT ATGAAAACCT  
 951 CCTGCTGCTT ATCGGCTCGG TATTGCGCC GATGGCGGCG GTTTTGATTG  
 1001 CCGACTTTTT CGTCTGAAA CGGCGTGAGG AGATTGAAGG CTTTGACTTT  
 1051 GCCGACTGG TTCTGTGGCT TCGGGGCTTC ATCCTCTACC GCTTCTGCT  
 1101 CTCGTCCGGC TGGGAAAGCA GCATCGGTCT GACCGCCCC GTAAATGTCTG  
 1151 CCGTGCCAT TGCCACCGTA TCGGTACGCC TTTTCTTTAA AAAAACCCAA  
 1201 TCTTACAAA GGAACCCGTC ATGA

10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

15  
 1 MSGNASSPSS SSAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KRGSVLFSA NMLQLAGWTA  
 101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT  
 151 VSMMLMLLAV LWLSAEVFST AGSTAAQVSD GMSFGTAVEL SAVMPLSWLP  
 201 LAADYTRHAR RPFATLTAT LAYTLTGCM YALGLAAALF TGETDVAKIL  
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGA SANNISARFA ETPVAVGVTL  
 301 IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEGFDF  
 351 AGLVLWLAGF ILYRELLSSG WESSIGLTAP VMSAVAIATV SVRLFVKKTQ  
 401 SLQRNPS\*

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

25  
 orf125.pep AGASANNISARFAETPVAVSVTLIGTVLAV  
 orf125a KILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAVVGTLLAV  
 250 260 270 280 290 300  
 30  
 orf125.pep MLPVTEYENFLLIGSVFAPMGGFDCRLFRLETAX  
 orf125a LLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG  
 310 320 330 340

35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

40  
 1 ATGTCGGGCA ATGCCTCCTC TCNTTCATCT TCCGCCGCCA TCGGGCTGAT  
 51 TTGGTTCGGC GCGCGCGTAT CGATTGCCGA AATCAGCAGC GGTACACTGC  
 101 TTGCGCCTTT GGGCTGGCAG CGCGCTCTGG CNGCTCTGCT TTTGGGTCAT  
 151 CCCGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC  
 201 CGGACNCANC TCGATGGAAA GCGTGCCTCT GTCGTTCCGC AAACGCGGTT  
 251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCCG CTGGACGGCG  
 301 GTGATGATT ACGCCGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT  
 351 GTGGGACGGC GAATCTTTTG TCTGGTGGGC ATTGGCAAAC GGCAGCTGA  
 401 TTGTGCTGTG GCTGTTTTC GGCGCACGCA AAACAGGCGG GCTGAAAACC  
 451 GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAANT  
 501 NTTTTCCACG GCAGGCAGCA CCGCCGANN GGTNNCAGAC GGCATGAGTT  
 551 TCGGAACGGC AGTCGAGCTG TCCGCCGTNA TGCCGCTTTC TTGGCTGCCG  
 601 CTGGCCGCCG ACTACACGCG CCACGCGCGC CGCCGTTTG CGGCAACCCT  
 651 GACGGCAACG CTCGCTTACA CGCTGACCGG CTGCTGGATG TATGCCCTGG  
 701 GTTTGGCAGC GCGTGTGTTT ACCGGAGAAA CCGACGTGGC AAAAATCCTG  
 751 CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTGC TCCTGTGCGAC  
 801 CGTTACCACC ACTTTTCTCG ATGCNTACTC CGCCGGCGTA AGTGCCAACA  
 851 ATATTTCGCG CAACTTTTCG GAAATACNA TCGCCGTTGC CGTCGCGGTT  
 901 CTCGGCACAC TGCTTGCCGT CCTCTGCCG GTTACCGAAT ATGAAAACCT  
 951 CCTGCTGCTT ATCGGCTCGG TATTGCGCC GATGGCGGCG GTTTTGATTG  
 1001 CCGACTTTTT CGTCTGAAA CGGCGTGAGG AGATTGAAGG C..

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

1 MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
 51 AVGGALFFAA AYIGALTGX SMESVRLSFG KRGSVLFSA NMLQLAGWTA

-444-

101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT  
 151 VSMLLMLLAV LWLSAEXFST AGSTAAXVXD GMSFGTAVEL SAVMPLSWLP  
 201 LAADYTRHAR RPFATLTAT LAYTLTGCWM YALGLAAALF TGETDVAKIL  
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPVAVAVAV  
 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	orf125a.pep	MSGNASSXSSSAAGLIWFGAAVSIAEISTGTLAPLWQORGLAALLGHAVGGALFFAA					
	orf125-1	MSGNASSPSSSSAAGLIWFGAAVSIAEISTGTLAPLWQORGLAALLGHAVGGALFFAA					
		10	20	30	40	50	60
15	orf125a.pep	AYIGALTGXXSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
	orf125-1	AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
		70	80	90	100	110	120
20	orf125a.pep	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEXFSTAGSTAAXVXD					
	orf125-1	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSAGSTAAXVXD					
		130	140	150	160	170	180
25	orf125a.pep	GMSFGTAVELSAVMPLSWPLAADYTRHARRPFAATLTATLAYTLTGCWMYALGLAAALF					
	orf125-1	GMSFGTAVELSAVMPLSWPLAADYTRHARRPFAATLTATLAYTLTGCWMYALGLAAALF					
		190	200	210	220	230	240
30	orf125a.pep	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAV					
	orf125-1	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVTL					
		250	260	270	280	290	300
35	orf125a.pep	VGTLLAVLLPVTEYENFLLLIGSVFAPMAAVLIADFFVLKRREEIEG					
	orf125-1	IGTVLAVMLPVTEYENFLLLIGSVFAPMAAVLIADFFVLKRREEIEGDFAGLVWLWLAGF					
		310	320	330	340	350	360

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETPVAVSVTLIGTVLAV	30
	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVTLIRTVLAV	308
50	orf125.pep	MLPVTEYENFLLLIGSVFAPM-GGFCRLFRLETA	64
	orf125ng	MLPVTEYKNFLLLIRSVFGPMAGGFCRLFCCLKTA	343

55 An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

	1	MSGNASSPSS	SAAIGLVWFG	AAVSIAEIST	GTLAPLWQ	RGLAALLGH
	51	AVGGALFFAA	AYIGALTGRS	SMESVRLSFG	KCGSVLFSVA	NMLQLAGWTA
	101	VMIYVATVS	SALGKVLWDG	ESFVWWALAN	GALIVLWLVF	GARRTGGLKT
	151	VSMLLMLLAV	LWLSVEVFAS	SGTNAAPAVS	DGMTFGTAVE	LSAVMPLSWL
60	201	PLAADYTRQA	RRPFAATLTA	TLAYTLTGCW	MYALGLAAL	FTGETDVAKI
	251	LLGAGLGTG	ILAVVLSTVT	TFLDTYSAG	ASANNISARF	AEIPVAVGVT
	301	LIRTVLAVML	PVTEYKNFLL	LIRSVFGPMA	GGFCRLFCCL	KTA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

      1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT
    51  TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
   101  TCGCCCCCTT GGGCTGGCAG CGCGGTCTGG CGGCCCTGCT TTTGGGTGTCAT
  151  GCCGTCGCGC GCGCGCTGTT TTTGCGGCG GCGTATATCG GCGCACTGAC
  201  CGGACGCAGC TCGATGGAAA GTGTGCGCCT GTCGTTCCGC AAATGCGGTT
  251  CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
  301  GTGATGATTT ACGTCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
  351  GTGGGACGGC GAATCCTTTG TCTGGTGGG ATTGGCAAAC GGCGCACTGA
  401  TCGTGCTGTG GCTGGTTTTT GCGCACGCA GAACGGGCGG GCTGAAAACC
  451  GTTTCGATGC TGCTGATGCT GCTTGCCGTG TTGTGGTTGA GCGTCGAAGT
  501  GTTCGCTTCG TCCGGCACA ACGCCGCGCC CGCCGTTTCA GACGGCATGA
  551  CTTTCGGAAC GGCAGTCGAA CTGTCCGCCG TCATGCCGCT TTCCTGGCTG
  601  CCGCTGGCCG CCGACTACAC GCGCCAAGCA CGCCGCCCGT TTGCGGCAAC
  651  CCTGACGGCA ACGCTCGCCT ATACGCTGAC GGGCTGCTGG ATGTATGCCT
  701  TGGGTTTGGC GCGGCTCTG TTTACCGGAG AAACCGACGT GGCGAAAATC
  751  CTGTTGGGCG CGGGCTTGGG CATAACGGGC ATTCTGGCAG TCGTCCTCTC
  801  CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCCGGC GCGAGTGCGA
  851  ACAACATTTT CGCGCGTTT GCGGAAATAC CCGTCGCTGT CGGCGTTACC
  901  CTGATCGGCA CGGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
  951  CTTCTGCTG CTTATCGGCT CGGTATTGCG GCCGATGGCG GCGGTTTTGA
 1001  TTGCCGACTT TTTCGTCTTA AAACGGCGTG AGGAGATGA AGGCTTTGAC
 1051  TTTGCCGGAC TGGTCTGTG GCTGGCAGGC TTCATCCTCT ACCGCTTCCT
 1101  GCTCTCGTCC GGTGGGAAA GCAGCATCGG TCTGACGCC CCGTAATGT
 1151  CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC
 1201  CAATCTTAC AAAGGAACCC GTCATGA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

      1  MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
    51  AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
   101  VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
   151  VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
   201  PLAADYTRQA RRPFAATLTA TLAYTLTGCW MYALGLAAAL FTGETDVAKI
   251  LLGAGLGITG ILAVVLSTVT TFFLDYSAG ASANNISARF AEIPVAVGVT
   301  LGTTLAVML PVTEYKNELL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
   351  FAGLVWLAG FILYRFLSS GWESSIGLTA PVMSAVAIAT VSVRLFKKKT
   401  QSLQRNPS*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

      10      20      30      40      50      60
  40  orf125-1.pep  MSGNASSPSSSSAIGLIWFGAAVVSIAEISTGTLLAPLGWQRGLAALLLGHAVGGALFFAA
      |||:|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     MSGNASSPSSSSAIGLVWFGAAVVSIAEISTGTLLAPLGWQRGLAALLLGHAVGGALFFAA
      10      20      30      40      50      60

      70      80      90      100     110     120
  45  orf125-1.pep  AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVAGATVSSALGKVLWDG
      70      80      90      100     110     120

      130     140     150     160     170     179
  50  orf125-1.pep  ESFVWWALANGALIVLWLVFGARTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAAQ-VS
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     ESFVWWALANGALIVLWLVFGARTGGLKTVSMLLMLLAVLWLSVEVFSSGTNAAPAVS
      130     140     150     160     170     180

      180     190     200     210     220     230     239
  55  orf125-1.pep  DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCWMYALGLAAAL
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCWMYALGLAAAL
      190     200     210     220     230     240

      240     250     260     270     280     290     299
  60  orf125-1.pep  FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVT
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf125ng-1     FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDYSAGASANNISARFAEIPVAVGVT
      250     260     270     280     290     300

```

-446-

		300	310	320	330	340	350	359
5	orf125-1.pep	LIGTVLAVMLPVTEYENFLL	LIGSVFAPMAAVLIADFFVL	KRREEIEGFD	FAGLVWLAG			
	orf125ng-1	LIGTVLAVMLPVTEYKNFLL	LIGSVFAPMAAVLIADFFVL	KRREEIEGFD	FAGLVWLAG			
		310	320	330	340	350	360	
10	orf125-1.pep	FILYRELLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
	orf125ng-1	FILYRELLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

20	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
25	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGT.AC GGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCCGAC	CGAAGC.CAG
	451	CTCGACGGGC	GGCAATTATA	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG..

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

	1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
	51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCRL	NTHTMQENG	SLIVWHGQDK
	101	PLSSEFVRHL	KRGXTDDEI	VRWRADDIAE	REPQLGGRFX	DGIYLPTEXQ
	151	LDGRQLXSAL	ADALDELNVP	CHWEHECVPE	ACK...	

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAGGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TTGCCGCCGC	CATGCTCGCG
40	151	CCTGCAGCGG	AAGCGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
45	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
	451	CTCGACGGGC	GGCAAAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GGCCTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCCGAGCA	CACCAGCACC	CTGCGCGGCA	TACGCGGCGA
	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCGTC
50	701	TGCTCCATCC	GCGTTATCCG	CTCTACATCG	CCCCGAAAGA	AAACCAACGTC
	751	TTCGTTCATCG	GCGCGACCCA	AATCGAAAGC	GAAAGCCAAG	CCCCCGCCAG
	801	CGTGCGTTCA	GGGTGGAAC	TCTTGTCGCG	ACTCTATGCC	ATCCACCCCG
	851	CTTTCGCGCA	AGCCGACATC	CTCGAAATCG	CCACCGGCCT	GCGCCCCACG
	901	CTCAACCAAC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCTGAT
55	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCCGCGGTAA
	1001	CCGCCGCCGC	CGCCAGATTG	GCAGTGGCAC	TGTTTGACGG	AAAAGACGCG
	1051	CCGGAACGCG	ATAAAGAAAG	CGGTTTGCGG	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEAVEATP EVVRLGRQSI PLWRGIRCL NTHTMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYGAKTA
201 WNQSPEHTST LRGIERGEVAR VYTPEITLNR PVRLLHPRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
301 LNHNNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAARL AVALFDGKDA
351 PERDKESGLA YIRRQD*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N.*

*meningitidis*:

```

15  orf126.pep      10      20      30      40      50      60
    orf126a      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAANLAPAAEVEATP
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAANLAPAAEVEATP
    10      20      30      40      50      60

20  orf126.pep      70      80      90      100     110     120
    orf126a      EVVRLGRQSIPLWRGIRCLNTHTMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    EVVRLGRQXIPLWRGIRCHLKTTPAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDXI
    70      80      90      100     110     120

25  orf126.pep      130     140     150     160     170     180
    orf126a      VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE
30  orf126a      130     140     150     160     170     180

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCNGGAA GGCTGACCGC
51  ACTGCAGCTT GCAGAACAAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
101 GCCGCCGGGG CGAACACGCC GCCGCCTATG TTGCCGCCGC CATGCTCGCG
151 CCTGCGGCGG AAGCGGTCGA AGCCACGCCT GAAGTGGTCA GGCTGGGCAG
201 GCAGANCATC CCGCTTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
251 CCATGATGCA NGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAA
301 CCTTTATCCA ACGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
351 TGACNAAATC GTCCGTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAAAATATT GTCTGCACTT GCCGACGCTT TGGACGAAC
501 GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCCGAA GACTTGCAAG
551 CCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
601 TGAACCAAT CCCCCGANNA NACCAGCACC CTGCGCGGCA TACGCGCGCA
45  651 AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGCTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGTC
751 TCGTCAATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGCGTTCC GGGCTGGAAC TCTTATCCGC ACTCTATGCC GTCCACCCCG
851 CCTTCGCGCA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG
50  901 CTCAATCACC ACAACCCGA AATCCGTTAC AACCGCGCCC GACGCCTGAT
951 TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA
1001 CCGCCGCCGC CGTCAGATTG CAGTGGCAC TGTGTGACGG AAAAGANGCG
1051 CCGAACGCG ATGAAGAAAG CGTTTGGCG TATATCCGAA GACAAGATTA
1101 A

```

This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEAVEATP EVVRLGRQXI PLWRGIRCHL KTPAMMXENG SLIVWHGQDK
101 PLSNEFVRHL KRGGVADDXI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAQYDWLI DCRGYGAKTA
201 WNQSPXXTST LRGIERGEVAR VYTPEITLNR PVRLLHPRYP LYIAPKENXV
60

```

-448-

251 FVIGATQIES ESCAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT  
 301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA  
 351 PERDEESGLA YIRRQD\*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5		10	20	30	40	50	60
	orf126a.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
	orf126-1	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
10		10	20	30	40	50	60
	orf126a.pep	70	80	90	100	110	120
	orf126-1	70	80	90	100	110	120
15		70	80	90	100	110	120
	orf126a.pep	130	140	150	160	170	180
	orf126-1	130	140	150	160	170	180
20		130	140	150	160	170	180
	orf126a.pep	190	200	210	220	230	240
	orf126-1	190	200	210	220	230	240
25		190	200	210	220	230	240
	orf126a.pep	250	260	270	280	290	300
	orf126-1	250	260	270	280	290	300
30		250	260	270	280	290	300
	orf126a.pep	310	320	330	340	350	360
	orf126-1	310	320	330	340	350	360
35		310	320	330	340	350	360
	orf126a.pep	310	320	330	340	350	360
	orf126-1	310	320	330	340	350	360
40		310	320	330	340	350	360
	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					
45		YIRRQDX					

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQGEHAAAYVAAAMLAPAAEAVEATP	60
	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
55	orf126ng	EVIRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE	180
	orf126ng	VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPO	180

60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTQGEHA AAYVAAAMLA

51 PAEEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK  
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA  
 201 WNQSPEHTST LRGIERGEVRG FTRPKSRSTA PCACCTRAIR STSPRKKTTS  
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRSM STPPSAKPTS SKWRPGLRPT  
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA  
 351 PERDEESGLA YIGRQD\*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC  
 51 ATTGACAGTT GCAGAACAAG GTTATCAGAT TGAACTTTTC GACAAGGGCA  
 101 CCCGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG  
 151 CCTGCGGCGG AAGCGGTCGA GGCAACGCCC GAAGTCATCA GGCTGGGCAG  
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA  
 251 CGATGATGCA GGAACACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG  
 301 CCATTATCCA GCGAGTTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA  
 351 TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC  
 401 AACTCGGCGG ACCTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG  
 451 CTCGACGGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT  
 501 GAACGTCCTT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG  
 551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAAACCGCG  
 601 TGGAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA  
 651 AGTGGCGCGG GTTTACACGC CCGAAATCAC GCTCAACCGC CCCGTGCGCG  
 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACACGTC  
 751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCAGCCAG  
 801 CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG  
 851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG  
 901 CTCACCCACC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCCTCAT  
 951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTTCC CCCGCCGTAA  
 1001 CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG  
 1051 CCGAACGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA  
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA  
 51 PAEEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK  
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA  
 201 WNQSPEHTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV  
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT  
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA  
 351 PERDEESGLA YIGRQD\*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

		10	20	30	40	50	60
45	orf126-1.pep	MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
	orf126ng-1	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAAEAVEATP					
		10	20	30	40	50	60
50	orf126-1.pep	EVVRLGRQSIPLWRGIRCRLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
	orf126ng-1	EVIRLGRQSIPLWRGIRCRLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
		70	80	90	100	110	120
55	orf126-1.pep	VRWRADDIAEREPQLGGRFS DGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE					
	orf126ng-1	VRWRADEIAEREPQLGGRFS DGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ					
		130	140	150	160	170	180
60	orf126-1.pep	GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP					
	orf126ng-1	DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP					
65		190	200	210	220	230	240

-450-

```

                250      260      270      280      290      300
orfl26-1.pep    LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAIHPAFGEADILEIATGLRPT
                |||||
orfl26ng-1      LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPT
                250      260      270      280      290      300

                310      320      330      340      350      360
orfl26-1.pep    LNHHNPEIRYNRRRLIEINGLFRHGFMI SPAVTA AARLAVALFDGKDAPERDKESGLA
                |||||
orfl26ng-1      LNHHNPEIRYSRERRLIEINGLFRHGFMI SPAVTA AAVRLAVALFDGKDAPERDEESGLA
                310      320      330      340      350      360

orfl26-1.pep    YIRRQDX
                || ||||
orfl26ng-1      YIGRQDX

```

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

```

gi|2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
Length = 327
Score = 169 bits (423), Expect = 3e-41
Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)

Query: 3   RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTRQGEHXXXXXXXXXXXXXXXXXXXXX 62
          RI V G G++G A QL G+++ L ++ G
Sbjct: 2   RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEPEV 60

Query: 63  IRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
          + LGR + W + G+L+V G+D F R G DE+
Sbjct: 61  LTLGRLAADWWEA-----LPGHVHRRGTLVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113

Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182
          IA EP L GRF ++ E LD RQ L+ALA L++ + +
Sbjct: 114 -----IAALEPDLAGRFRRLFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165

Query: 183 QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLLLHPRYPY 242
          +D V+DC G LRG+RGE+ V T E++L+RPVRLLLHPR+P+Y
Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETTEVSLSRPVRLLLHPRHPY 218

Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
          I P++ + F++GAT IES+ P + RS +ELL+A YA+HPAFGEA + E AG+RP
Sbjct: 219 IVPDRKNRFMVGATMIESDDGGPITARSLMELLNAAAYAMHPAFGEARVTETGAGVRPAYP 278

Query: 303 HHNPEIRYSRERRLIEINGLFRHGFMI SP 331
          + P R ++E R + +NGL+RHGF+++P
Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLLAP 305

```

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

```

50 819>:
      1  ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
      51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
     101  TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAATGCA
     151  CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGGTTTA AACAAACATC
     201  TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
     251  GTTTGAATGG AATCGtCGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG
     301  AAGGCGGTAG CCATAGATAA AGATAAAAAT CCTTTTATTA TTAAGATGAA
     351  TGAAAATCTA GTAACCTTTA aTTTGCAAGA AGTCCGCCAG TTCGTGTAGT
     401  GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
     451  GTAG

```



-451-

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

      1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
    51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
   101 KAVIDKDKN PFIKMENLV VTFICKKSAS SCSDGLDYFK GNDKCKLLK
   151 *

```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

      1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
    51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
   101 TTGAGAAAGC AAAGATAAAT GCAGTGC GGGAGGTTTA AGAAAATGCA
   151 CATTTTATGG AAAAGTTTTC TCTGCAGAA GGGAGGTTTA AACAAACATC
   201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
   251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
   301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
   351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
   401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

      1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
    51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
   101 AVAIDKDKNP FIIKMENLV TFICKKSASS CSDGLDYFKG NDKCKLLK*

```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N. meningitidis*:

```

25      orf127.pep      10      20      30      40      50      60
      MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENHFMEKFYLN
      |||||||
      orf127a      MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLENHFMEKFYLN
      10      20      30      40      50      60

30      orf127.pep      70      80      90      100      110      120
      GRFKQTSTKWPSLPIKEAEGFCIRLNGIVARXALDSKFMLKAVAIDKDKNPFIKMENLV
      |||||||
      orf127a      GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFIKMENLV
      70      80      90      100      110

35      orf127.pep      130      140      150
      VTFICKKSASSCSDGLDYFKGNDKCKLLKX
      |||||||
      orf127a      VTFICKKSASSCSDGLDYFKGNDKCKLLKX
40      120      130      140      150

```

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

      1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
    51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
   101 TTGAGAAAGC AAAGATAAAT ACAGTGC GGGAGGTTTA AGAAAATGCA
   151 CATTTTATGG AAAAGTTTTC TCTGCAGAA GGGAGATTTC AACAAACATC
   201 TACCAAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
   251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
   301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
   351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
   50 401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

      1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN
    51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
   101 AVAIDKDKNP FIIKMENLV TFICKKSASS CSDGLDYFKG NDKCKLLK*

```

-452-

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

```

      10      20      30      40      50      60
orfl27a.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLEN AHFMEKFYLQN
5 orfl27-1   MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl27a.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
10 orfl27-1   GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
      70      80      90     100     110     120

      130     140     150
orfl27a.pep TFICKKSASSCSDGLDYFKGNDKDKCKLLKX
15 orfl27-1   TFICKKSASSCSDGLDYFKGNDKDKCKLLKX
      130     140     150

```

## 20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

```

orfl27.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN 60
25 orfl27ng MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAAFLENA HFMEKFYLQN 60

orfl27.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARXALDSKFMLKAVAIDKDKNPFI IKMNENL 120
30 orfl27ng GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFI IKMNENL 119

orfl27.pep VTFICKKSASSCSDGLDYFKGNDKDKCKLLK 150
orfl27ng VTFICKKSASSCSDRLDYFKGNDKDKCKLLK 149

```

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

```

35 1 ATGACTGATA ATCGGGGGTT TACTACTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGATTTA AACAAACATC
201 TACCAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
40 251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 826>:

```

45 1 MTDNRGFTLV ELISVVLILSVLALIVYPSY RNYVEKAKIN AVRAAFLENA
51 HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIIKMENLV TFICKKSASS CSDRLDYFKG NDKDKCKLLK*

```

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

```

      10      20      30      40      50      60
50 orfl27-1.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN
orfl27ng-1 MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN
      10      20      30      40      50      60

      70      80      90     100     110     120
55 orfl27-1.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
orfl27ng-1 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
      70      80      90     100     110     120
60

```

-453-

orf127-1.pep      130      140      150  
                  TFICKKSASSCS DGLDYFKGNDKDCKLLKX  
                  |||||  
 orf127ng-1      TFICKKSASSCS DGLDYFKGNDKDCKLLKX  
                  130      140      150

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

1    ..GTGTCGCTGG CTTCGGTGAT TGCCCTCTCAA ATCTTCCTTT ACGAAGATTT  
 51    CAACCAAATG CGGAAAACCC GTGGAGCTAT CTGCGGTTTT CTTGTCCAAT  
 101    ATTTATCTGG GGTTCAGCA GGGGTATTTC GATTTGAGTG CCGACGAGAA  
 151    CCCCCTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC  
 201    TGTATCCCTT TTTGCTGATA TTTTGCTGCA AAAAAACCAA ATCGCTACGG  
 251    GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC  
 301    GTTTTGGCCA AGCGGGTTT ATACCGACAT CCTCAACCAA CCCAATACTT  
 351    ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTGGCAGG TTCGCTGCTG  
 401    GCGGTTTACG GGCAACGCA AAACGGCAGA CGGCAACAG CAAATGGAAA  
 451    ACGGCAGTTC CTTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT  
 501    TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT  
 551    CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT  
 601    TCCGACCCGC ATCCTGTCCG CAAGCCCCAT CGTATTGTGC GGCAAAATCT  
 651    CTTATTCCTT ATACCTGTAC CATTGGATT TTATTGCTTT CGCTCCGCTC  
 701    ATTAGAGGCG GGAACAGCT CGGACTGCCT GCCG..

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

1    ..VSLASVIASQ IFLYEDFNQM RKTVELSAVF LSNIYLGFFQ GYFDLSADEN  
 51    PVLHIWSLAV EEQYLLYPL LLIFCCKKTK SLRVLRNISI ILFLILTASS  
 101    FLPSGFYTDI LNPNTYYLS TLRFPPELLAG SLLAVYGGTQ NGRRTQANGK  
 151    RQLSSSLCFG ALLACLFVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL  
 201    PTRILSASPI VEVGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

1    ATGCAAGCTG TCCGATACAG ACCGAAATT GACGGATTGC GGGCCGTCGC  
 35    CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG  
 101    GATTCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTAAC  
 151    GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT  
 201    TTATACCCGC AGGATTAAGC GGATTATCC TGCCTTTATT GCGGCCGTGT  
 251    CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTC AAC  
 40    CAAATGCGGA AAACCGTGGA GCTTCTGCG GTTTCTTGT CCAATATTTA  
 351    TCTGGGGTTT CAGCAGGGGT ATTCGATTT GAGTGCCGAC GAGAACCCCG  
 401    TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT  
 451    CCCCTTTTGC TGATATTTTG CTGCAAAAAA ACCAAATCGC TACGGGTGCT  
 501    GCGTAACATC AGCATCATCC TGTTTGTGAT TTTGACTGCC TCATCGTTTT  
 45    TGCCAAGCGG GTTTTATACC GACATCCTCA ACCAACCCAA TACTTATTAC  
 601    CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTCGC TGCTGGCGGT  
 651    TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAT GGAAAACGGC  
 701    AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCCCTG CCTGTTCTGT  
 751    ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCCTG  
 801    CCTGCTGACG GCACTGCTTA TCCGGAGTAT GCAATACGGG ACACTTCCGA  
 851    CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTGGGCAA AATCTCTTAT  
 901    TCCCTATACC TGTACCATTT GATTTTATT GCTTTCGCC ATTACATTAC  
 951    AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA  
 1001    CGGCCGATT TTCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA  
 55    AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTCT ATCTCGCCCC  
 1101    GTCCCTGATA CTTGTGCGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC  
 1151    AGGAACACCT CCGCCCGTTG CCCGCGCGC CCCTTGCTGC GGAAATCAT

10

15 This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

20

25

Computer analysis of this amino acid sequence gave the following results:

30 Homology with hypothetical integral membrane protein HI0392 of *H.influenzae* (accession number U32723)

ORF128 and HI0392 show 52% aa identity in 180aa overlap:

35

40

Homology with a predicted ORF from *N.meningitidis* (strain A)

45 ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N. meningitidis*:

```

                                10      20      30
orf128.pep                      VSLASVIASQIFLYEDFNQMRTVELSAVF
50 orf128a    ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAIVLASVIASQIFLYEDFNQMRTVELSAVF
              60       70       80       90      100     110
                                40      50      60      70      80      90
orf128.pep    LSNIYLGFQQGYFDLSADENPVLHIWISLAVEEQYYLLYPLLLIFCCKTKTSLRVLRNISI
60 orf128a    LSNIYLGFQQGYFDLSADENPVLHIWISLAVEEQYYLLYPLLLIFCCKTKTSLRVLRNISI
              120     130     140     150     160     170
                                100     110     120     130     140     150
orf128.pep    IILFLILTASSFLPSGFTDILNQPNYYLSTIRFPELLAGSLLAVYGQTONGRRQTANGK
              :      :      :      :      :      :

```

-455-

orf128a		ILFLILTATSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	180	190	200	210	220	230
5	orf128.pep	RQLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	160	170	180	190	200	210
	orf128a	RQLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	240	250	260	270	280	290
10	orf128.pep	VFVGKISYSLYLHWIFIAFAPLIRGGKQLGLPA	220	230	240			
	orf128a	VFVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR	300	310	320	330	340	350
15	orf128a	KMTFFKKAFFCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSH	360	370	380	390	400	410

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

20	1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTCGC
	51	CGTGCTATCC	GTCATGATT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
	101	GATTCTCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
	151	GGCATCATT	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTT	TCCGGGATT
25	201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	TGCTTTTATT	GCGGCCGTGT
	251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
	301	CAATGCGGA	AAACCGTGA	GCTTCTGCG	GTTTTCTGT	CCAATATTTA
	351	TCTGGGGTT	CAGCAGGGT	ATTTCGATT	GAGTGCCGAC	GAGAACCCG
30	401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
	451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAATCGC	TACGGGTGCT
	501	GCGTAACATC	AGCATCATCC	TATTTCTGAT	TTTGACTGCC	ACATCGTTTT
	551	TGCCAAGCGG	GTTTATATAC	GATATTCTCA	ACCAACCCAA	TACTTATTAC
35	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTCCG	TGCTGGCGGT
	651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAACCGGC
	701	AGTTGCTTTC	ATCACTCTGC	TTCCGGCGCAT	TGCTTGCCCTG	CCTGTTCCGT
	751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCCTGC	TCCTTCCCTG
40	801	CCTGCTGACG	GCACTGCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
	851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
	901	TCCCTATACC	TGTACCATTG	GATTTTATT	GCTTTCGCCC	ATTACATTAC
	951	AGCGGACAAA	CAGCTCGGAC	TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
45	1001	CGGCCGGATT	TTCCCTGTTG	AGTTATTATT	TGATTGAACA	GCCGCTTAGA
	1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATT	TTCTGCCTCT	ATCTCGCCCC
	1101	GTCCCTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAAC
	1151	AGGAACACCT	CCGCCCGTTG	CCCGGCGCGC	CCCTTGCTGC	GGAAAATCAT
50	1201	TTTCCGAAA	CCGTCCTGAC	CCTCGGCGAC	TCGCACGCCG	GACACCTGCG
	1251	GGGGTTTCTG	GATTATGTCG	GCAGCCGGGA	AGGGTGGAAG	GCCAAAATCC
	1301	TGTCCTTCGA	TTCCGAGTGT	TTGGTTGGG	TAGATGAGAA	GCTGGCAGAC
	1351	AACCGTTAT	GTCGAAAATA	CCGGGATGAA	GTTGAAAAAG	CCGAAGCCGT
55	1401	TTTCATTGCC	CAATCTATG	ATTTGAGGAT	GGGCGGCCAG	CCCGTGCCGA
	1451	GATTTGAAGC	GCAATCCTTC	CTAATACCCG	GGTCCCAGC	CCGATTCAGG
	1501	GAACCGTCA	AAAGGATAGC	CGCCGTCAA	CCCGTCTATG	TTTTTGCAA
	1551	CAACACATCA	ATCAGCCGTT	CGCCCTGAG	GGAGGAAAAA	TTGAAAAGAT
60	1601	TTGCCGCAA	CCAATATCTC	CGCCCATTC	AGGCTATGGG	CGACATCGGC
	1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGATT	AAAGATATTC	CCAATGTGCA
	1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATACG
	1751	GCCGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGGTTCTTAT
65	1801	TATATGGGGC	GGGAATTTCA	CAAACACGAA	CGCCTGCTTA	AATCTTCTCG
	1851	CGACGGCGCA	TTGCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 832>:

60	1	MQAVRYRPEI	DGLRAVAVLS	VMIFHLNNRW	LPGGFLGVDI	FFVISGFLIT
	51	GIILSEIQNG	SFSRDEYTR	RIKRIYPAFI	AAVSLASVIA	SQIFLYEDFN
	101	QMRKTVELSA	VFLSNIIYLG	QQGYFDLSAD	ENPVLHIWSL	AVEEQYLLLY
	151	PLLIIFCCKK	TKSLRVLRNI	SIILFLILTA	TSFLPSGFYT	DILNQPNNTYY
65	201	LSTLRFPPELL	AGSLLAVYGQ	TQNGRRQTAN	GKRQLSSSLC	FGALLACLFV
	251	IDKHNPFIPG	MTLLLPCLLT	ALLIRSMQYG	TLPTIRLSAS	PIFVVGKISY
	301	SLYLYHWIFI	AFAHYITGDK	QLGLPAVSAV	AALTAGFSLL	SYYLIEQPLR
	351	KRKMTFFKAF	FCLYLAPSLI	LVGYNLYARG	ILKQEHRLPL	PGAPLAAENH
70	401	FPETVLTGLD	SHAGHLRGFL	DYVGSREGWK	AKILSLDSEC	LVWVDEKLAD
	451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	PVPRFEAQSF	LIPGFARFR

501 ETVKRIAARK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG  
 551 KSNQAVFDLI KDIPNVHWVD AQKYLKNTV EIYGRYLYGD QDHLTYFGSY  
 601 YMGREFHKHE RLLKSSRDGA LQ\*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDDIFFVISGFLITGIILSEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDDIFFVISGFLITGIILSEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128a.pep	QQGYFDLSADENPVLHIWISLAVEEQYYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
15	orf128-1	QQGYFDLSADENPVLHIWISLAVEEQYYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRLSSLC
	orf128-1	SSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRLSSLC
20	orf128a.pep	FGALLACLFVIDKHNPFIPIGMTLLLPCLLTALLIRSMQYGTLPTRILSASPISVFGKISY
	orf128-1	FGALLACLFVIDKHNPFIPIGMTLLLPCLLTALLIRSMQYGTLPTRILSASPISVFGKISY
25	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
30	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHRLRPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHRLRPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
35	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSFILIPGFAPRFETVKRIAARKPVYVFANNTSISRSPLREEKLKRFAANQYL
	orf128-1	PVPRFEAQSFILIPGFAPRFETVKRIAARKPVYVFANNTSISRSPLREEKLKRFAANQYL
40	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKIDIPNVHWVDAQKYLKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKIDIPNVHWVDAQKYLKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSHGALQX

#### Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
55	orf128.pep	LSNIYLGFGQGYFDLSADENPVLHIWISLAVEEQYYLLYPLLLIFCCKKTKSLRVLNISI	90
	orf128ng	LSNIYLGFRGFDLSADENPVLHIWISLAVEEQYYLLYPLLLIFCYKKTKSLRVLNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGK	232
	orf128.pep	RQLSSSLCFGALLACLFVIDKHNPFIPIGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
65	orf128ng	RQLSSSLCFGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292

```

orf128.pep  VFVVGKISYSLYLHWIFIAFAPLIRGGKQLGLPA      244
             |||||
orf128ng    VFVVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR  352

```

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

```

1  ATGCAAGCTG TCCGATACAG GCCTGAAATT GACGGATTGC GGGCCGTCGC
51  CGTGCTATCC GTCATTATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101 GATTCCCTGGG GGTGGACATT TTCTTTGTCA TCTCGGGATT CCTCATTACC
151 AACATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
10  201 TTATACCCGC AGGATTAAGC GGATTATACC TGCTTTTATT GCGGCCGTGT
251 CCTGGCTTC GGTGATTGCT TCTCAAATCT TCCTTACGA AGATTCAAC
301 CAAATGAGGA AAACCATAGA GCTTCTACG GTTTTTTGT CCAATATTTA
351 TTTGGGGTTC CGATTGGGGT ATTTGATTT GAGTGCCGAC GAGAACCCCG
401 TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT
15  451 CCTCTTTTGC TGATATTCTG TTACAAAAAA ACCAAATCAC TACGGGTGCT
501 GCGTAATATC AGCATCATCC TGTTCTGAT TTTGACCGCA TCATCGTTTT
551 TGCCGGCCGG GTTTTATACC GACATCCTCA ACCAACCCaa TACTATTAC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTCG TGTGGCGGT
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGAAAAT GGAAAACGGC
20  701 AGTTGCTTTC ATTACTCTGT TTCGGCGCat tgCTGTCTG CCTGTCTGTG
751 ATCGACAAAC ACGATCCGTT TATCCCGGGA ATAACCTGC TCCTTCCCTG
801 CCTGCTGACG GCGCTGCTTA TCCGGAGTAT GCAATACGGG ACACTTCCGA
851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTG GATTTTATT GCCTTCGCCC ATTACATTAC
25  951 AGGCGACAAA CAGCTCGGAC TGCTGCGGT ATCGGCGGT GCCGCGTTGA
1001 CGGCCGGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTAGA
1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTTT ATCTCGCCCC
1101 GTCCCTGATG CTTGTGCGTT ACAACCTGTA TTCAAGAGGG ATATTGAAAC
1151 AGGAACACCT CCGCCCGCTG CCCGGCACGC CCGTTGCTGC GGAAAAAAT
30  1201 TTTCCGGAAG CCGTCTTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG
1251 GGGGTTTCTG GATTATGTCG GCGGCAGGGA AGGGTGGAAG GCTAAAATCC
1301 TGTCCTCTGA TTCGAGTGT TTGGTTTGGG TGGATGAGAA GCTGGCAGAC
1351 AACCCTTGT GCCGAAAATA CCGGATGAA GTTGAAAAAG CCGAAGCTGT
40  1401 TTTGATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA
1451 GATTTGAAGC GCAATCCTTC CTGATACCCG GGTCAAAGC CCGATTCAGG
1501 GAAACCGTCA AGAGGATAGC CGCCGTCAAA CCTGTATATG TTTTTCGAAA
1551 CAATACATCA ATCAGCCGTT CTCCTTGAG GGAGGAAAAA TTGAAAAGAT
1601 TTGCTATAAA CCAATACCTC CGGCCTATTC GGGCTATGGG CGACATCGGC
40  1651 AAGAGCAATC AGGCGGTCTT TGATTGGTT AAAGATATTC CCAATGTGCA
1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG
1751 GACGCTATCT TTACGGCGAC CAAGACCACC TGACCTATT CCGTTCTTAT
1801 TATATGGGGC GGAATTTCA CAAACACGAA CGCCTGCTCA AGCATTCCCC
1851 AGGCGGCGCA TTGCAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 834>:

```

45  1  MQAVRYRPEI DGLRAVAVLS VIIFHLNLRW LPGGFLGVDI FFVISGFLIT
51  NIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTIELST VFLSNIYLG FRLGYFDLSAD ENPVLHIWSL AVEEQYYLLY
151 PLLLIFCYKK TKSLRVLRLNI SIILFLILTA SSELPA GFYT DILNQPN TTY
50  201 LSTLRFPELL VGSLLAVY GQTQNGRRQ TEN GKRQLSLLC FGALLVCLFV
251 IDKHDPFIPG ITLLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSL SYYLIEQPLR
351 KRKMTFFKAF FCLYLAPSLM LVGYNLYSRG ILKQEHRLPL PGTPVAAENN
401 FPETVLTLDG SHAGHLRGFL DYVGGREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFKA RFR
55  501 ETVKRIA AVK PVYVFANNTS ISRSPLREEK LKRFAINQYL RPIRAMGDIG
551 KSNQAVFDLV KDIPNVHWVD AQKYLKNTV EIHGRLYLDG QDHLTYFGSY
601 YMGREFHKHE RLLKHSRGGA LQ*

```

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

```

60  orf128-1.pep  MQAVRYRPEIDGLRAVAVLSVMIFHLNLRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
             |||||
orf128ng      MQAVRYRPEIDGLRAVAVLSVIIFHLNLRWLPGGFLGVDIFFVISGFLITNIILSEIQNG

orf128-1.pep  SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
             |||||
65  orf128ng      SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG

```

-458-

```

5  orf128-1.pep  QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA
   orf128ng     RLGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCYKKTSLRVLNRNISIILFLILTA

10 orf128-1.pep  SSFLPSGFYTDILNQPNNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSLLC
   orf128ng     SSFLPAGFYTDILNQPNNTYYLSTLRFPELLVGSLLAVYGQTQNGRRQTENGKRQLLSLLC

15 orf128-1.pep  FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY
   orf128ng     FGALLVCLFVIDKHDFFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY

20 orf128-1.pep  SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFFKAF
   orf128ng     SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFFKAF

25 orf128-1.pep  FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLDSSHAGHLRGFL
   orf128ng     FCLYLAPSLMLVGYNLYSRGILKQEHRLPLPGTPVAAENNFETVLTGLDSSHAGHLRGFL

30 orf128-1.pep  DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
   orf128ng     DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ

35 orf128-1.pep  PVRFEAQSFILPGFPARFRETVKRIA AVKPVYVFANNTSISRSPREEKLRFAANQYL
   orf128ng     PVRFEAQSFILPGFKARFRETVKRIA AVKPVYVFANNTSISRSPREEKLRFAINQYL

40 orf128-1.pep  RPIQAMGDIGKSNQAVFDLIKDI PNHVWVDAQKYLEPKNTVEIYGRYLYGDQDHLTYFGSY
   orf128ng     RPIRAMGDIGKSNQAVFDLVKDI PNHVWVDAQKYLEPKNTVEIHGRYLYGDQDHLTYFGSY

45 orf128-1.pep  YMGREFHKHERLLKSSHGGALQX
   orf128ng     YMGREFHKHERLLKHSRGGALQX
                        610      620

```

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

```

40 sp|P43993|Y392_HAEIN  HYPOTHETICAL PROTEIN HI0392 >gi|1074385|pir||B64007
   hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
   >gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
   influenzae] Length = 245
   Score = 239 bits (604), Expect = 3e-62
   Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)

45 Query: 38 VDIFFVISGFLITNIILSEIQNSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97
   +DIFFVISGFLIT II++EIQ SFS + FYTRRIKRIYP F+Y
   Sbjct: 1 MDIFFVISGFLITGIIITEIQNSFSFLKQFYTRRIKRIYPAFITVMALVSFIASAIFIYN 60

50 Query: 98 DFNQMRKTIELSTVFLSNIYLGFRGLGYFDLSADENPVLHIWSLAVEEQXXXXXXXXXIFC 157
   DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWSLAVE Q I
   Sbjct: 61 DFNKLRKTIELAIAFLSNFYLGLTQGYFDLSANENPVLHIWSLAVEGQYYLIFPLILILA 120

55 Query: 158 YKKTSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPELLVGSLLAV 217
   YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+
   Sbjct: 121 YKKFREVKVLFIIITLILFILLATSFVSANFYKEVLHQPNIIYYLSNLRPELLVGSLLAI 180

60 Query: 218 YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDFFIPGIT 262
   Y N + Q +L++L L CLF+++ + FIPGIT
   Sbjct: 181 YHNLN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

```

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.



**Example 99**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51 GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATG GGTCTGTTGT
101 TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCCGATGCCG
151 GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201 CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTGGGGCG TATGTGTGGT
251 TTCCGTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51 VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTTC GTTTTGACAT TATTTACGAA TACCGCTGGA TGTTTCTTTA
51 CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
151 CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
151 GCGCGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
201 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501 GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCCG TCATTGCTGT GGCGGAGTTG
251 GCGTATGTTT AGAATACGAT TACGGGCCCG TATTGCGTTT ATGAAGAACC
601 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701 GGATATTCTT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

1  MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIWIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDDSS LLSVIAVAEL
201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

```

40 orf129.pep      10      20      30      40      50
      IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
orf129a      MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
              10      20      30      40      50      60

45 orf129.pep      60      70      80
      ALRKVSLLYVTLFRGTPLEFVQIWIWAYVWFPPFFV
orf129a      ALRKVSLLYVTLFRGTPLEFVQIWIWAYVWFPPFFVHPSDGI
              70      80      90      100     110     120

50 orf129a      SLALIANSQAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS
              130     140     150     160     170     180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTTC GTTTTGACAT TATTTACGAA TACCGCTGGA TGTTTCTTTA
51 CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT

```

-460-

101 CCGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA  
 151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT  
 201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA  
 251 TTTGGGCGTA TGTGTGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT  
 5 301 TTGGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT  
 351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG  
 401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG  
 451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT  
 10 501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA  
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATGCTGT GCGCGAGTTG  
 601 GCGTATGTTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC  
 651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT  
 701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

This encodes a protein having amino acid sequence <SEQ ID 840>:

15 1 MDFRFDIIYE YRWMFLYAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
 51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVHPSDGI  
 101 LVSGEAAIAL RRGYGPIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA  
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL  
 201 AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR\*

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

orf129a.pep MDFRFDIIYEYRWMFLYALTTGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 orf129-1 MDFRFDIIYEYRWMFLYALTTGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 25 orf129a.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPIAG  
 orf129-1 ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPIAG  
 30 orf129a.pep SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS  
 orf129-1 SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS  
 orf129a.pep EFITLLKDSS LLSVIAVAEL AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE  
 35 orf129-1 EFITLLKDSS LLSVIAVAEL AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE  
 orf129a.pep KRYNPQHRX  
 orf129-1 KRYNPQHRX  
 40

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

45 orf129.pep IIEYRWMFLYALTTGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW 54  
 orf129ng MDFRFDIIYEYRWMFLYALTTGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 60  
 orf129.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFV 88  
 50 orf129ng ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVILHTAFLGNAMRQSRRVPDKGRWIAG 120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

1 MDFRFDIIYE YRWMFLYAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
 55 51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVILHTAF  
 101 LGNAMRQSRR VPDKGRWIAG SLELNCQPRG RKTRGEFPFG ESNLGTPEPN  
 151 PLSMGQRRFP GCENWYPPQN FIKK\*

Further work revealed the following gonococcal sequence <SEQ ID 843>:

1 ATGGATTTTc gtTTTGACAT TATTTAcgaA TACCGCTGGA TGTTTCTTTA

-461-

51 CGGCGCACTG Acgaccttgg ggctgacggt cgtggcgacg gCGGGCGGTT  
 101 CGGtattggG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA  
 151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT  
 201 GCTGTACGTT ACCCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA  
 251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT  
 301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT  
 351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG  
 401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG  
 451 GCGTGTCTT TGGGACTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT  
 501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA  
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCCG TCATTGCTGT GGCGGAGTTG  
 601 GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGTTT ATGAAGAACC  
 651 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL  
 701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
 51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVHPSDGI  
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA  
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL  
 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR\*

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

orf129-1.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 25 orf129ng-1 MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 orf129-1.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 orf129ng-1 ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
 30 orf129-1.pep SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 orf129ng-1 SLALIANSQA YICEIFRAGI QSIDKGQMEA ACSLGLTYPQ AMRYVILPQA LRRMLPLAS  
 35 orf129-1.pep EFITLLKDSS LLSVIAVAEL AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 orf129ng-1 EFITLLKDSS LLSVIAVAEL AYVQNTITGR YSVYEEPLYT AALIYLLMTT FLGWIFLRLE  
 orf129-1.pep KRYNPQHRX  
 ||||||||  
 40 orf129ng-1 KRYNPQHRX

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)  
 [Archaeoglobus fulgidus] Length = 224  
 Score = 132 bits (329), Expect = 2e-30  
 45 Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)  
 Query: 65 VSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAGSLAL 124  
 +S YV + RGTP L VQI + I + F P + GI + E A G + AL  
 50 Sbjct: 58 ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99  
 Query: 125 IANSQA YICEIFRAGI QSIDKGQMEA ACSLGLTYPQ AMRYVILPQA LRRMLPLASEFIT 184  
 SQA YI EI RAGI + SI QMEA SLG + TY QAMRYVI PQA R + LP L + EFI  
 Sbjct: 100 SICSA YIAEIVRAGIESIPIGQMEA ARSLGMTYLQAMRYVIFPQAFRNILPALGNEFIA 159  
 55 Query: 185 LLKDSS LLSVIAVAEL AYVQNTITGR YSVYEEPLYT AALIYLLMTT FLGWIFLRLEKR 242  
 LLKDSS LLSVI ++ EL V I P AL YL + MT L + + K +  
 Sbjct: 160 LLKDSS LLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQKK 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests  
 that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful  
 60 antigens for vaccines or diagnostics, or for raising antibodies.

**Example 100**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1   ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
51  TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
101 GGCTGCCCCG GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
151 CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACCTT TACGTAAACA
201 cTACGTCCGC ACTTATTACy TGCTCCAACT CTTTGCCGCC GCAGgcTAgT
251 TTGTGGACAG GCGCGGCGwA ATTACAAAAC CTGCCCGCyT CCGCGCCCTT
301 GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
10  TGACCGCCCG ACTGTGGCAC AGCGGCTTTA CCAAACCTGA CTACCCCAAA
401 CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
451 TTTCTTGGrTG AACGTGAACC CGrTATTTTt CATTACCGTT CcTGCGATTc
501 TGACCGCCGC CGTATTTCGT CTGTATCTTT TCrCGTTTAT ACCGATATTT
551 CGGCGGAATG CGTTTACAGA CGATCCGGAr TAr

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1   ..LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPAAPL
101 HLITLGGMMG GVMVWLTAAG LWHSGFTKLD YPKLCRIAPV ILFAAAVSRA
151 FLXNVNPXFF ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1   ATGCGGCCGT TTTTCGTCGG CGCGCGGGTG CTTGCCATAC TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101 TGGAACTTAT GCTGCGGCG GCATACGGCG GTTTTGTGAC TGCGGCTTTG
151 TTGGACTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTGC CGACTTTGAT
201 GCGCGCATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAAA
251 CTGCCCTCGT TTTCTGCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
301 GCCCGGCTGA TTTGGCTAGA CCGAAACACC GACAACTTCG CCCTGCTAAT
351 GTTACTTGCC GCGTTCACCTG TTTTTCAGAC GGCATATGCC GTCAGCGGCG
401 ATTTGAACCT GTTGCAGCGC CAAGTGCATC TAAATATGGC GGCGGTGATG
10  451 TTCGTATCCG TGCGCGTCAG TATTCTTTTG GGCGCGGAAG CCCTGAAAGA
501 ATGCCGCTCTG AAAGACCCCTG TTTTATTTC AAATATCGTT TATAAAAAA
551 TCGCCATTAC TTTCTGCTC TTGCAGCCCG CCGCGGAAC TTGGCTGCCC
601 GCGCAAACCG CCGGTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
651 CAAGCTGCGT GAGCTTCACC ATCAGAACT CTTAGCTAAA CACTACGTCC
10  701 GCACTTATTA CCGCTCCAA CTCTTGCCG CCGCAGGCTA TTTGTGGACA
751 GGCGCGGCGA AATTACAAAA CCTGCCCGCC TCCGCGCCCC TGCACCTGAT
801 TACCTCGGC GGCATGATGG GCGGCGTGAT GATGGTGTGG CTGACCGCCG
851 GACTGTGGCA CAGCGGCTTT ACCAACTCG ACTACCCAA ACTCTGCCGC
901 ATTGCCGTCC CCATCCTTT CGCGCGCGC GTCTCGCGC CTTTCTTGAT
10  951 GAACGTGAAC CCGATATTT TCATTACCGT TCCTGCGATT CTGACCGCCG
1001 CCGTATTCGT ACTGTATCTT TTCACGTTA TACCGATATT TCGGGCGAAT
1051 GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1   MRPFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
45  51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVL LLLFC
101  ARLIWLD RNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
151  FVSVRVS ILL GAEALKECRL KDPVFIPNIV YKNIAITELL LHAAELWLFP
201  AQTAGFTA LA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAAAGYLWT
251  GAAKLQNL PA SAPHLITLG GMMGGVMMVW LTAGLWHS GF TKLDYPKLCR
50  301  IAVPILFAA A VSRAFLMNVN PIFFITVPAI LTAAEVLYL FTFIPIFRAN
351  AFTDDE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

-463-

```

                    10      20      30
orf130.pep          LKECRLKDPVFIPNIVYKNIATFLLHAA
                    |||||
orf130a             LNLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNVYKNIATFLLHAA
                    140      150      160      170      180      190

                    40      50      60      70      80      90
orf130.pep          AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX
                    |||||
orf130a             AELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX
                    200      210      220      230      240      250

                    100      110      120      130      140      150
orf130.pep          LQNLPAAPLHLITLGGMMGGVMMVWLTAGLWHSFGFTKLDYFKLCRIAVPILFAAAVSRA
                    |||||
orf130a             LQNLPAAPLHLITLGGMMGSVMVWLTAGLWHSFGFTKLDYFKLCRIAVPILFAAAVSRA
                    260      270      280      290      300      310

                    160      170      180      190
orf130.pep          FLXNVNPFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPEX
                    |||||
orf130a             VLMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPEX
                    320      330      340      350

```

The complete length ORF130a nucleotide sequence <SEQ ID 849> is:

```

25      1  ATGCGGCCGT  TTTTCGTCGG  CGCGCGCGTG  CTTGCCATAC  TCGGTGCGCT
      51  GGTGTTTTTC  ATCAACCCCG  GTGCCATCGT  CCTGCACCGC  CAAATTTTCT
     101  TGGAACTTAT  GCTGCGGCGC  GCATACGGCG  GTTTTTTGAC  TCGCGCTTTG
     151  TTGGACTGGA  CGGGTTTTTC  GGGTAACCTG  AAACCTGTCG  CGACTTTGAT
     201  GCGCGCATTA  TTGCTCGCCG  CATCCGCTAT  ACTGCCCTTT  TCGCCGCAAA
     251  CTGCCCTGTT  TTTCTGTCGC  GCCTATTGGC  TGGTGTGCT  GCTGTCTGCT
     301  CCCCAGCTGA  TTTGGCTAGA  CCGAAACACC  GACAACTTCG  CCCTGCTAAT
     351  GTTACTTGCC  GCGTTCACTG  TTTTTCAGAC  GGCATATGCC  GTCAGCGCGC
     401  ATTTGAACCT  GTTGCAGCGC  CAAGTGCATC  TAAATATGGC  GCGCGTGATG
     451  TTCGTATCCG  TCGCGCTCAG  TATTCTTTTG  GCGCGGAAG  CCCTGAAAGA
     501  ATGCCGCTCG  AAAGACCCAG  TATTCATCCC  CAATGTCGTC  TATAAAAAACA
     551  TCGCCATTAC  CTTCTGCTC  CTGCAGCCG  CCGCCGAAC  TTGGCTGCCT
     601  GCGCAAACCG  CCGGTTTTAC  CTCGCTCGCC  GTCGGCTTTA  TCCTGCTTGC
     651  CAAGCTGCGT  GAGCTTCACC  ATCAGCAACT  CCTGCGCAA  CACTACGTCC
     701  GCACTTATTA  CCTGCTCCAA  CTCTTTGCCG  CCGCAGGCTA  TTTGTGGACA
     751  GCGCGGCGCA  AATTACAAAA  CCTGCCCGCC  TCCGCGCCCC  TGCACCTGAT
     801  TACCCTCGGT  GGCATGATGG  GCAGCGTGAT  GATGGTGTGG  CTGACTGCCG
     851  GACTGTGGCA  CAGCGCTTT  ACCAAGCTCG  ACTACCGAA  ACTCTGCCGC
     901  ATCGCCGTCC  CCATCCTNTT  CGCCGCGCC  GTTTCGCGCG  CTGTTTTAAT
     951  GAACGTAAAC  CCGATATTCT  TCATACCCGT  CCCCACAATT  CTGACCGCCG
    1001  CCGTGTTCGT  GCTTTACCTG  CTGACATTCG  TACCGATCTT  TCGGGCGAAC
    1051  GCGTTTACAG  ACGATCCGGA  ATAA

```

This encodes a protein having amino acid sequence <SEQ ID 850>:

```

      1  MRPFFVGA AV  LAILGALVFF  INPGAIVLHR  QIFLELM LPA  AYGGFLTAAL
     51  LDWTGFSGNL  KPVATLMAAL  LLAASAILPF  SPQTASFFVA  AYWLVL L LFC
    101  ARLIWLD RNT  DNFALLM LLA  AFTVFQTA YA  VSGDLNLLRA  QVHLNMAAVM
    151  FVSVRVSILL  GAEALKECRL  KDPVFIPNVV  YKNIAITFLL  LHAAELWLP
    201  AQTAGFTSLA  VGFI LLAKLR  ELHHHELLRK  HYVRTYYLLQ  LFAAAGYLWT
    251  GAAKLQNLPA  SAPLHLITLG  GMMGSVMVW  LTAGLWHSGF  TKLDYFKLCR
    301  IAVPILFAAA  VSRV LNMVN  PIFFITVP AI  LTA AVFVLYL  LTFVPIFRAN
    351  AFTDDPE*

```

ORF130a and ORF130-1 show 98.3% identity in 357 aa overlap:

```

orf130a.pep        MRPFFVGA AV  LAILGALVFF  INPGAIVLHR  QIFLELM LPA  AYGGFLTAAL LDWTGFSGNL
|||
orf130-1            MRPFFVGA AV  LAILGALVFF  INPGAIVLHR  QIFLELM LPA  AYGGFLTAAL LDWTGFSGNL

orf130a.pep        KPVATLMAALLAASAILPFSPQTASFFVAAYWLVL L LFCARLIWLD RNTDNFALLM LLA
|||
orf130-1            KPVATLMAALLAASAILPFSPQTASFFVAAYWLVL L LFCARLIWLD RNTDNFALLM LLA

orf130a.pep        AFTVFQTA YAVSGDLNLLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNVV

```

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

20	orf130.pep	LKECRLKDPVFIPNIVYKNIAITFLLLHAA	30
	orf130ng	LNNLLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRLKDPVFIPNVIYKNIAIT-LLHAA	201
25	orf130.pep	AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX	90
	orf130ng	AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAX	261
30	orf130.pep	LQNLPASAPLHLITLGGMMGCVMMVWLTAGLWHSGFTKLDYPKLCRIAVPILFAAAVSRA	150
	orf130ng	LQNLPASAPLHLITLGGMTGGVMMVWLTAGLWHSGFTKLDYPKLCRIAVSILFASAVSRA	321
	orf130.pep	FLXNVNXPFFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPE	193
	orf130ng	VLMNVNPIFFITVPEILTAAVFMYLLTTFVPIFRANAFTDDPE	364

35 An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

40

1	MNKFFTHPMR	PFFVGAAVLA	ILGALVFFHQ	PRRYHPAPPN	FLGTYAAGCI
51	RREFDYREVG	PDGFFRQPET	CRYFDGGVVA	CCGCFIAVFT	ATCRIFERRL
101	LAGVA AVLRL	ADLARRQHRT	LRSVDVTAAE	TVFQTAYAVS	GDNLRLRAQV
151	HLNMAVMFV	SVRVSVLLGT	ETLKECRLLD	PVPFIPNVIYK	NIAITLLLHV
201	AAELWLPAQT	AGFTALAVGF	ILLAKLRELH	HHELLRKHYV	RTYYLLQLFA
251	AAGYLWTGAA	KIQNLSPASAP	LHLITLGGMT	GGVMVWLTA	GLWHSGETKL
301	DYPKLCRIAV	SILFASAVSR	AVLMNVNPIF	FITVPEILTA	AVFMLYLLTF
351	VPIFRANAFT	DDPE*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

45	1	ATGCGCCCGT	TTTTCGTCGG	TGCGGCAGTA	CTTGCCATAC	TCGGTGC GTT
	51	GGTGT TTTT T	ATCAACCCCG	GC GCTATCAT	CCTGCACCGC	CAAATTTTCT
	101	TGGAAC TTA	GCTGCCGGCT	GCATACGGCG	GTTTTTTGAC	TACCGCTTTG
	151	TTGGACCGGA	CGGGTTTTTC	AGGCAACCTG	AAACCTGCCG	CTACTTTGAT
	201	GGCGGTGTTG	TTGCTTGTTG	CGGCTGTTTT	ATTGCCGTTT	TTACCGCAAC
50	251	TTGCCGCA TT	TTTCGTGCGC	GCCTATTGGC	TGGTGTGTCT	GCTGTTCTGC
	301	GCCTGGCTGA	TTTGGCTCGA	CCGCAACACC	GACAAC TTCG	CTCTGTTGAT
	351	GTTACTTGCC	GCATTTACCG	TTTTTTCAGC	GGCCTATGCC	GTCAGCGCGC
	401	ATTTGAAC TT	ACTGCGCGCG	CAAGTGCATT	TGAATATGGC	GGCGGT CATG
	451	TTCGTATCCG	TC CGCGTCAG	CGTCTCTTTG	GGCACGGAAA	CCCTGAAAGA
55	501	ATGCCGTCCT	AAAGACCCCG	TATTCATCCC	GACGTTATTC	TATAAAAACA
	551	TCGCCATCAC	CCTGCTGCTG	CACGCCCGCG	CCGAAC TTTG	GCTGCCCGCG
	601	CAAACCGCCG	GTTTTACTGC	GCTTGCCGTC	GGCTTCATCC	TGCTCGCCAA
	651	GCTGCCGCAA	CTGCACCATC	ACGAAC TCTT	ACGCAAAAC	TACGTC CGCA
	701	CTTATTACCT	GCTCCAGCTC	TTTGCCGCGG	CAGGTTATCT	GTGGACAGGC
60	751	GCGGCGAAAC	TGCAAAACCT	GCCCGCCTCC	GCGCCCTCG	ACCTGATTAC
	801	CCTCGCGCGC	ATGACGGGTG	CGGTGATGAT	GGTGTGGCTG	ACTGCCCGAC
	851	TGTGGCACAG	CGCGCTTACC	AAACTCGACT	ACCGCAAACT	CTGCCGCTAC

```

901 GCCGTCTCCA TCCTTTTCGC CTCGCCGCTT TCGCGCGCTG TTTTAATGAA
951 CGTGAATCCG ATATTCTTCA TCACCGTTCC CGAGATTCTG ACCGCCGCCG
1001 TGTTTCATGCT TTACTGCTG ACGTTCGTAC CGATTTTTCG AGCGAACGCC
1051 TTTACAGACG ATCCGGAATA A

```

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

```

1 MRPFFVGA AV LAILGALVFF INPGAII LHR QIFLELM LPA AYGGFLTTAL
51 LDRTGFGS NL KPAATLMAVL LLVA AVLLPF LPQLAAFFVA AYWLVL LLLFC
101 AWLIWLD RNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAAE LWLPA
120 201 QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYLLQL FAAAGYLWTG
251 AAKLQNL PAS APLHLITLGG MTGGVMMVWL TAGLWHS GFT KLDYPKLCRI
301 AVSILFASAV SRAVLMNVNP IFFITVPEIL TAAVFM LYL TFVPIFRANA
351 FTDDPE*

```

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

```

15 orf130-1.pep MRPFFVGA AVLAILGALVFFINPGAIVLHRQIFLELM LPAAYGGFLTAALLDWTGFGS NL
orf130ng-1 MRPFFVGA AVLAILGALVFFINPGAII LHRQIFLELM LPAAYGGFLTTALLDRTGFGS NL

20 orf130-1.pep KPVATLMAALLAASAILPFSPQTASFFVAAYWLVL LFCARLIWLD RNTDNFALLMLLA
orf130ng-1 KPAATLMAV LLLVA AVLLPFLPQLAAFFVAAYWLVL LFCAWLIWLD RNTDNFALLMLLA

orf130-1.pep AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSI LLGAEALKECRLKDPVFIPNIV
25 orf130ng-1 AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRLKDPVFIPNVI

orf130-1.pep YKNIAITFLLHAAAE LWLPAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYLLQ
orf130ng-1 YKNIAIT-LLLHAAAE LWLPAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYLLQ

30 orf130-1.pep LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMTGGVMMVWL TAGLWHS GFTKLDYPKLCR
orf130ng-1 LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMTGGVMMVWL TAGLWHS GFTKLDYPKLCR

35 orf130-1.pep IAVPILFAAAVSRAFLMNVNPIFFITVPAILTA AVFVLYLFTFIPIFRANAFTDDPEX
orf130ng-1 IAVSILFASAVSRAVLMNVNPIFFITVPEILTA AVFM LYLTFVPIFRANAFTDDPEX

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

```

1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCA TT
51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
45 101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT
151 GCGGCGGAGA GTCCGCCGTC TTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA
351 CTGCTTGGA AAG..

```

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

```

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51 GGESPPSLGD YEIPLSDGNS SVRANEYESA QSYFYRKIG KFEXCGLDWR
101 TRDGKPLIET FKQGGFDCLE K..

```

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

```

55 1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCA TT
51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

```

-466-

101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TGGGATATT  
 151 GGGCGCGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA  
 201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
 251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT  
 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA  
 351 CTGCTTGGAA AAGCAGGGGT TGC GCGCAA CGGTCTGTCC GAGCGCGTCC  
 401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI  
 51 GGESFPPSLGD YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR  
 101 TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf131.pep		MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESFPPSLGD					
orf131a		MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESFPPSLED					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf131.pep		YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE					
orf131a		YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK					
		70	80	90	100	110	120
orf131.pep	K						
orf131a	KQGLRRNGLSERVRWX						
	130						

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

1 ATGGAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT  
 35 51 TACGTTGCA GGCTGCCGCT TGGCAGGTTG GTATGAGTGT TCGTCCCTGT  
 101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATT TGGGATATT  
 151 GGGCGCGAGA GTCCGCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA  
 201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
 251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGT GGATTGGCGT  
 40 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTTGA  
 351 TTGTTTGAAG AAGCAGGGGT TGC GCGCAA CGGTCTGTCC GAGCGCGTCC  
 401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI  
 45 51 GGESFPPSLED YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR  
 101 TRDGKPLIET FKQEGFDCLK KQGLRRNGLS ERVRW\*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

orf131a.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESFPPSLED
orf131-1	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESFPPSLGD
orf131a.pep	YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK
orf131-1	YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE
orf131a.pep	KQGLRRNGLSERVRWX



-467-

orf131-1 KQGLRRNGLSERVRWX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

```

orf131.pep  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  60
            ||||:||||| |||:||||||| |||:||||||| |||:||||||| |||:||||||| |||
orf131ng    MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED  60

10 orf131.pep  YEIPLSDGNSSVRANEYESAQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE  120
            ||||| ||| ||||| |||:||||||| |||:||||||| |||:||||||| |||
orf131ng    YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE  120

15 orf131.pep  K 121
            |
orf131ng    KQGLRRNGLSERVRW 134

```

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

```

20      1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
      51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
     101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

Further work revealed the following gonococcal DNA sequence &lt;SEQ ID 863&gt;:

```

25      1 ATGGAAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTCGATT
      51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCTTGT
     101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATT TTTGGGATATT
     151 GCGGCGGAGA GtcgctGTC TTAGAGGAC TACGAGATAC CGCTTTCAGA
     201 CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAAATCTT
     251 ACTTTTATAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
     301 ACGCGTGACG GCAAACCTTT GGTGAGAGG TTCAAACAGG AAGGTTTCGA
     351 CTGTTTGAA AAGCAGGGGT TCGGCGCAA CGGCCTGTCC GAGCGCGTCC
     401 GATGGTAA

```

This corresponds to the amino acid sequence &lt;SEQ ID 864; ORF131ng-1&gt;:

```

35      1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
      51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
     101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

40 orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED
            ||||:||||| |||:||||||| |||:||||||| |||:||||||| |||:||||||| |||
orf131-1        MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD

orf131ng-1.pep  YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
            ||||| ||| ||||| |||:||||||| |||:||||||| |||:||||||| |||
orf131-1        YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

45 orf131ng-1.pep KQGLRRNGLSERVRWX
            ||||| ||| ||||| |||
orf131-1        KQGLRRNGLSERVRWX

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 102**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

      1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
      51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
5      101  AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
      151  TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
      201  CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
      251  TGAACCTCGG CCTGCCTTAT ATtTcCGGCC CGCAATGGCT GTCGGAAGAAAC
      301  GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
10     351  GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCTCGCGC
      401  CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATtTCGGCG TTTCCGCCCG
      451  CCTGCCGCAA ACGCCGCGCC AAGACCGGAA CAGCCAATCG CCGTTTtTcG
      501  TCATCGAAGC CGACGAATAC GACACCGCCT TtTCGACAA ACGTTCTAA
15     551  TtCGTGcATT ACCGTCCGCG TACCGCGGTG TTGAACAATC TGGAAATCGA
      601  CCACGCCGAC ATCTTTGCCG ACTTGGGCGC GATACAGACc CAGTTCCACT
      651  ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
      701  CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGA
      751  AAAATTCCGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

20     1  MKHIHIIGIG GTFMGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
      51  YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
      101  VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKFRFRFP
      151  PAANAAPRPE QPIAVFRHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
25     201  PRRHLCLRLR DTDVPLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAGG
      251  KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

      1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
      51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
30     101  AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
      151  TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
      201  CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
      251  TGAACCTCGG CCTGCCTTAT ATTTCCGGCC CGCAATGGCT GTCGGAAGAAAC
      301  GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
35     351  GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
      401  CGGGCTTCCT TATTGGCGGC GTACCGGAAA ATTTCCGGCGT TTCCGCCCGC
      451  CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
      501  CATCGAAGCC GACGAATACG ACACCGCCTT TTTTCGACAA CGTTCTAAAT
      551  TCGTGcATTA CCGTCCGCGT ACCGCGGTGT TGAACAATCT GGAATTCGAC
40     601  CACGCCGACA TCTTTGCCGA CTTGGGCGCG ATACAGACCC AGTTCCACTA
      651  CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
      701  AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA
      751  AAATTTCGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCACGG
45     801  CTCGTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
      851  ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTGCCGCC
      901  GCGCGTCATG TCGGTGTCGA TATTCAGACC GCCTGCGAAG CCTTGGGCGC
      951  GTTTAAAAAC GTCAAACGCC GGATGGAAAT CAAAGGCACG GCAAACGGCA
50     1001  TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
      1051  ATTCAGGTTT TCGGCCAACG CGTCCGCGGC GCGCGCATCC TCGCCGTCCT
      1101  CGAACC CGCT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
      1151  CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGCGCTG
      1201  GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
      1251  CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
      1301  TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
      1351  GGAAAGCTGC TGGAAAGCTTT GAGATAG

```

55 This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

      1  MKHIHIIGIG GTFMGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
      51  YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
      101  VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
60     151  LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
      201  HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTFVE
      251  KFGTEHGWA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
      301  ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITYYDDF AHHPTAIETT

```

-469-

351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLKEA DQVFCYAGGV  
 401 DWDVAEALAP LGGRNLVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH  
 451 GKILLEALR\*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E. coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

10 Orf132: 4 IHIIGIGGTFMGGLAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAQAQLDEFK 63  
 IHI+GI GTFMGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +  
 o457: 3 IHILGICGTFMGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q 61  
 Orf132: 64 ADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTASML 123  
 D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTTA M  
 o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA 121  
 15 Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143  
 W+LE G PGF+IGGV G  
 o457: 122 TWILEQCGYKPGFVIGGVPG 141

Homology with a predicted ORF from *N. meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

25 orf132.pep 10 20 30 40 50 60  
 MKHIHIIGIGGTFMGGLAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAQAQLD  
 orf132a 10 20 30 40 50 60  
 MKHIHIIGIGGTFMGGIAIAKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD  
 30 orf132.pep 70 80 90 100 110 120  
 EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA  
 orf132a 70 80 90 100 110 120  
 EFKADVYVIGNVAKRGMDVVEAILNRLGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTTA  
 35 orf132.pep 130 140 150 160  
 SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQFI-----AVFR  
 orf132a 130 140 150 160  
 SMLAWVLEYAGLAPGFXIGGVPEFNSVSARL-PQTPRQDPNSQSFFVIEADEYDTAFFD  
 40 orf132.pep 170 180 190 200 210 220  
 HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHLCLGRDTPVPLPRAYRAVXRL  
 orf132a 180 190 200 210 220 230  
 KRSKFVHYRPTAVLNNLEFDHADIFADLGAIQTQFHHLVTRVPSEGLIVCNGRQQLQD

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

1 ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGTGGGAT  
 51 TGCCGCCATT GCCAAAGAAG CAGGGTTTGA ANTCAGCGGT TGCGATGCGA  
 101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTG  
 151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTTAAAG CCGACGTTTA  
 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT  
 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC  
 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC  
 351 GACCACCGCG TCTATGCTCG CGTGGGTTT GGAATATGCC GGACTCGCAC  
 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC  
 451 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT  
 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAT  
 551 TCGTGCAATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC  
 601 CACGCCGACA TCTTCGCCGA TTGGGGCGCG ATACAGACCC AGTTCACCA  
 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACCGC  
 701 AGCAAAGCCT GCAAGACACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA  
 751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGATGG

-470-

801 CTCGTTTCGAC GTGTTGCTTG ACGGCAAAAA AGCCGGACAC GTCGCTTGGA  
 851 GTTTGATGGG CGGACACAAC CGCATGAACG CGCTCGCNGT CATCGCCGCC  
 901 GCGCGTCATG CCGGAGTNGA CATTCAGACG GCCTGCGAAG CCTTGAGCAC  
 951 GTTTAAAAAC GTCAAACGCC GCATGGAAT CAAAGGCACG GCAAACGGTA  
 1001 TCACCGTTTA CGACGACTTC GCCCACCATC CGACCGCTAT CGAAACCACG  
 1051 ATTCAAGGTT TGCGCCAGCG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT  
 1101 CGAACCGCGT TCCAATACGA TGAAGCTGGG TACGATGAAA GCCGCCCTGC  
 1151 CCGCAAGCCT CAAAGAAGCC GACCAAGTGT TCTGNTACGC CGCGCGCGCG  
 1201 KFTGGGACG TTGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGCACGT  
 1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG  
 1301 CAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC  
 1351 ACCAAACTGC TGGACGCTTT GAGATAG

This encodes a protein having amino acid sequence <SEQ ID 870>:

1 MKHIHIIGIG GTFMGGIAAI AKEAGFEXSG CDAKMYPPMS TQLEALGIGV  
 51 YEGFDTAQLD EFKADVYVIG NVAKRGMDEV EAILNRGLPY ISGPQWLAEN  
 101 XLHHHWXLGV AXTHGKTTTA SMLAWVLEYA GLAPGFXIGG VPENFSVSAR  
 151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD  
 201 HADIFADLGA IQTQFHHLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTPVE  
 251 KFGTEHWQA GEANADGSFD VLLDGKKAGH VAWSLMGGHN RMNALAVIAA  
 301 ARHAGVDIQT ACEALSTFKN VKRMEIKGT ANGITVYDDF AHHPTAIETT  
 351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK AALPASLKEA DQVFXYAGGA  
 401 DWDVAEALAP LGGRLHVGKD FADFVAEIVK NAEAGDHILV MSNGGFGGIH  
 451 TKLLDALR\*

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

25 orf132a.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD  
 orf132-1 MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAAQLD  
 30 orf132a.pep EFKADVYVIGNVAKRGMDEV EAILNRGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTTA  
 orf132-1 EFKADVYVIGNVAKRGMDEV EAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA  
 35 orf132a.pep SMLAWVLEYAGLAPGFXIGGVFENFSVSARLPQTTPRQDPNSQSPFFVIEADEYDTAFFDK  
 orf132-1 SMLAWVLEYAGLAPGFLIGGVFENFGVSARLPQTTPRQDPNSQSPFFVIEADEYDTAFFDK  
 40 orf132a.pep RSKFVHYRPR TAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGRQQSLQDT  
 orf132-1 RSKFVHYRPR TAVLNNLEFDHADIFADLGAIQTQFHYLVRTVPSEGLIVCNGRQQSLQDT  
 45 orf132a.pep LDKGCWTPVEKFGTEHWQAGEANADGSFDVLLDGKKAGHVAWSLMGGHNRMNALAVIAA  
 orf132-1 LDKGCWTPVEKFGTEHWQAGEANADGSFDVLLDGKTAGRVKDWLMGRHNRNALAVIAA  
 50 orf132a.pep ARHAGVDIQTACEALSTFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIIQGLRQRVGG  
 orf132-1 ARHVGVDIQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIIQGLRQRVGG  
 55 orf132a.pep ARILAVLEPR SNTMKLGTMKAALPASLKEADQVFXYAGGADWDVAEALAPLGGRLHVGKD  
 orf132-1 ARILAVLEPR SNTMKLGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNVGKD  
 orf132a.pep FADFVAEIVKNAEAGDHILVMSNGGFGGIHKTLLDALRX  
 orf132-1 FADFVAEIVKNAEAGDHILVMSNGGFGGIHGLLEALRX

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

60 orf132.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAAQLD 60  
 orf132ng MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDKMYPPMSTQLEALGIGVHEGFDAAQLE 60

-471-

	orf132.pep	EFKADVYVIGNVAKRGMVDVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMVDVEAILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTA	120
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRRFRPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLRRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHLRRLGRDTPVPPFRAHRTIRRHRLQRTAAKPARY	240
	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
15	orf132ng	FGQRLLDAGGKIRHRLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
20	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDDV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWLVGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPKFRFRFRP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
	201	PRRHLRRLGR	DTDPVPPRA	HRTIRRHRL	QRTAAKPARY	FGQRLLDAGG
	251	KIRHRLAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CCGGGTCAA	AGTCAGCGGT	TGCGACGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGGCT	TGCATGCCGC	GCACTTGCAA	GAATTTCAAG	CCGATATTTA
30	201	CGTCATCGGC	AATGTCCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATTT
	251	TGAACCGTGG	GCTGCCCTTAT	ATTTCCGCEC	CGCAATGGCT	GGCTGAAAac
	301	GTGCTgcacc	atcaTTGGgt	ACTCGGCGTG	GcagggacGC	ACGGcaaaAc
	351	gaccaCcGcg	tCCATGCTCG	CCTGGGTCTT	GGAATATGCC	GGACTCGCGC
	401	CGGGCTTCCT	CATCGGCGGT	gtaccggaAA	ATTTCCGGCT	TTCCGCCCCG
35	451	CTACCGCAAA	CGCCGCGTCA	AGACCCGAAC	AGCAAATCGC	CGTTTTTCGT
	501	CATCGAAGCC	GACGAATACG	ACACCGCCTT	TTTCGACAAA	CGCTCCAAAT
	551	TGCTGCATTA	TGCCCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCTTCGCCGA	CTTGGGCGCG	ATACAGACCC	AGTTCCACCA
	651	CCTCGTGCGC	ACCGTACCAT	CCGAAGGCCT	CATCGTCTGC	AACGGACAGC
	701	AGCAAAGCCT	GCAAGATACT	TTGGACAAAG	GCTGCTGGAC	GCCGGTGGAA
40	751	AAATTTCGCA	CCGGACACGG	CTGGCAGATT	GGTGAAGTCA	ATGCCGACGG
	801	CTCGTTTCGAC	GTATTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTCGCATGGG
	851	ATTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCCGT	CATCGCTGCC
	901	GCACGCCATG	CCGGAGTCGA	TGTTTCAGACG	GCCTGCGAAG	CCTTGGGTGC
45	951	GTTTAAAAAC	GTCAAACGCC	GCAATGGAAAT	CAAAGGCACG	GCAAACGCGA
	1001	TCACCGTTTA	CGACGATTTT	GCCCACCACC	CGACCGCCAT	CGAAACCACG
	1051	ATTCAAGGTT	TGCGCCAACG	TGTCGGCGGC	GCGCGCATCC	TCGCCGTCCT
	1101	CGAGCCGCGT	TCCAACACCA	TGAAACTCGG	CACGATGAAG	TCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGCTACGC	CGGCGGCGCG
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCTGCA	GGCTGCGCGT
50	1251	CGGTAAAGAT	TTCGATACCT	TCGTTGCCGA	AATTGTGAAA	AACGCCCGAA
	1301	CCGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACTGC	TGGACGCTTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
55	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDDV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWLVGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPENFGVSAR
	151	LPQTTPRQDPN	SKSPFFVIEA	DEYDTAFFDK	RSKFVHYRPR	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGQQQSLQDT	LDKGCWTPVE
	251	KFGTGHGWQI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGGHN	RMNALAVIAA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRRMEIKGT	ANGITVYDDF	AHHPTAIEET
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DQVFCYAGGA
	401	DWDVAEALAP	LGRLRVGKD	FDTFVAEIVK	NARTGDHILV	MSNGGFGGIH
	451	TKLLDALR*				

-472-

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5  orf132ng-1.pep MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLE
   orf132-1      MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD
10 orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA
   orf132-1      EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA
15 orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK
   orf132-1      SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK
20 orf132ng-1.pep RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVTRTPSEGLIVCNGQQQSLQDT
   orf132-1      RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHYLVTRTPSEGLIVCNGRQQSLQDT
25 orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSFDVLLDGKKAGHVAWDLMGHNRNMNALAVIAA
   orf132-1      LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNMNALAVIAA
30 orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
   orf132-1      ARHVGVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
35 orf132ng-1.pep ARILAVLEPRSNMKGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGRLRVGKD
   orf132-1      ARILAVLEPRSNMKGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGRLRVGKD
40 orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
   orf132-1      FDAFVAEIVKNAEVDHILVMSNGGFGGIHGKLLLEALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E.coli* protein:

```

35  pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
   ORF_o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
   in fbp-pmba intergenic region [Escherichia coli] Length = 457
   Score = 474 bits (1207), Expect = e-133
   Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)
40  Query: 22 KEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
   ++ G +V+G DA +YPPMST LE GI + +G+DA+QLE Q D+ +IGN RG VE
   Sbjct: 21 RQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79
45  Query: 82 AILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141
   A+L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV
   Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCGYKPGFVIGGV 139
50  Query: 142 PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRTAVLNNLEFDH 201
   P NF VSA L +S FVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH
   Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190
55  Query: 202 ADIFADLGAIQTQFHHLVTRTPSEGLIVCNGQQQSLQDTLDDKGCWTPVEKFGTGHGWQIG 261
   ADIF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ
   Sbjct: 191 ADIFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQELVGEQGHWAQAK 250
60  Query: 262 EVNADGS-FDVLLDGKKAGHVAWDLMGHNRNMNALAVIAAARHAGVDVQTACEALGAFKN 320
   ++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N
   Sbjct: 251 KLTTDASEWEVLLDGEKVGKSLVGEHNHNGLMIAAARHVGVPADAANALGSFIN 310
65  Query: 321 VKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNMKGTM 379
   +RR+E++G ANG+TVYDDFAHHPTAI T+ LR +VGG ARI+AVLEPRSNMKG+G
   Sbjct: 311 ARRRLELRGEANGVTYDDFAHHPTAILATLAALRGKVGGTARIIAVLEPRSNMKG+GIC 370
70  Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGRLRVGKDFDTFVAEIVKNARTGDHI 438
   K L SL AD+VF W VAE D DT +VK A+ GDHI
   Sbjct: 371 KDDLAPSLGRADEVFLQPAHIPWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI 430
   Query: 439 LVMSNGGFGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L  
 Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These  
 10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

### Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```

1   ..CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
51  CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
15  101  CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
    151  ATTAGTGCGG ACTTCGGCGA TTATTTTCATG CCGTTCGCCA GCTATTCGCG
    201  CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTC CAAATCGGCG
    251  ACTCCGGCGT TCACACCGCC TAAAACCAAG AGCGCGCAAA CACTTGGCAA
    301  TTTGGCTTCr ATACCTATAA AAAAGGATTG TAAAACAAG ATGATACATT
20  351  AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA
    401  ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTC GAGCTGGGTC
    451  AGCAGCACCG GGCTTGCTTA CACCATCCAA CATCGCrATT TCaWAGACAA
    501  AGTGCATCAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGTT
    551  TTTTCACCAA CCTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCAAC
25  601  TTCAGCGATG CGAGCGAATC GCCCAACAAT CCGTCCAAG AAGACCAACT
    651  CAAACAAGGT TATGGGTTGA GCAGGGTTTC CGCCCTGCCG CGAGATTACG
    701  GAGCTTTTGA ACTCGGTACG CGCTGGTTGG GCAACAACCT GACTTTGGGC
    751  GCGCGGATGC GCTATTTTCG CAAGAGCATC CGCGCGACGG CTGAAGAACG
    801  CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTC CGGCAACTGG
30  851  GCAAGCGTTC CATCAACAA ACCGAACTC TTGCCCGCCA GCCTTTGATT
    901  TTWgATTTTa ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTCCGCGC
    951  CGAAGTCAA AATCTGTTTC ACAGGCGTTA TATCGATCCG CTCGATGCGG
    1001 GCAATGATGC GGCAAC.GAG CGTTATTACA GCTCGTTTCA CCCGAAAGAC
    1051 AAGGACrrAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA
35  1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
    1151 CCTTTTgAT GACGATGAGC TACAAGTTTT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```

1   ..PGYYGSDEF KRAFGENSPT XKKHCHNSCG IYEPVLKKYG KKRANNHSVS
40  51  ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
    101  FGFXTYKKGL LKQDDTLGLK LVGYRSRIDN YIHNVYGKWW DLNGDIPSWV
    151  SSTGLAYTIQ HRXFXDKVHQ XXXXXXXXYD YGRFFTNLSY AYQKSTQPTN
    201  FSDASESPNN ASKEDQLKQG YGLSRVSALP RDYGRLEVGT RWLGNKLTIG
    251  GAMRYFGKSI RATAEERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI
    301  XDFNAAYEPK KNLIrFAEVK NLFDRRYIDP LDAGNDAAAXE RYYSFDPKD
45  351  KDXDVTcNAD KTLcNGKYGG TSKSVLTNFA RGRTFLMTMS YKF*
  
```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```

1   GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
51  ACCGAAAGAC AAAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC
50  101  AGGATATATT CAAATCCAGC GAAAACCTCG ACAACATCGT ACGCAGCATC
    151  CCCGGTGCGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA
    201  TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGGACGGCA
    251  TCACGCAGAC CTTTTATTTCG ACTTCTACCG ATCGGGCGAG GGCAGGCGGT
  
```

-474-

301 TCATCTCAAT TCGGTGCATC TGTGACAGC AATTTTATG CCGGACTGGA  
 351 TGTCGTCAAA GGCAGCTTCA GCGGCTCGGC AGGCATCAAC AGCCTTGCCG  
 401 GTTCGGCGAA TCTGCGGACT TTAGGCGTGG ATGACGTCGT TCAGGGCAAT  
 451 AATACCTACG GCCTGCTGCT AAAAGSTCTG ACCGGCACCA ATTCAACCAA  
 501 AGGTAATGCG ATGGCGGCGA TAGGTGCGCG CAAATGGCTG GAAAGCGGAG  
 551 CATCTGTCGG TGTGCTTTAC GGGCACAGCA GGGCGAGCGT GGGCGAAAAT  
 601 TACGCGGTGG GCGGCGGCGG GCAGCACATC GGAATTTTGG GCGCGGAATA  
 651 TTTGGAACGG CGCAAGCAGC GATATTTTGT ACAAGAGGGT GCTTTGAAAT  
 701 TCAATTCGGA CAGCGGAAAA TGGGAGCGGG ATTTACAAAG GCAACAGTGG  
 751 AAATACAAGC CGTATAAAAA TTACAACAAC CAAGAATAC AaAAATACAT  
 801 CGAAGAGCAT GACAAAAGCT GCGGGGAAAA CCTg . CaCCG CAATACGACA  
 851 TTACCCCAT CGATCCGTCC AGCCTGAAGC AGCAGTCGGC AGGCAATCTG  
 901 TTTAAATTGG AATACGACGG CGTATTCAAT AAATACACGG CGCAATTTCTG  
 951 CGATTTAAAC ACCAAAATCG GCAGCCGCAA AATCATCAAC CGCAATTTATC  
 1001 AGTTCAATTA CGGTTTGTCT TTGAACCCGT ATACCAACCT CAATCTGACC  
 1051 GCAGCCTACA ATTGCGGCAG GCAGAAATAT CCGAAAGGGT CGAAGTTTAC  
 1101 AGGCTGGGGG CTTTAAAGG ATTTTGAAC CTACAACAAC GCGAAAATCC  
 1151 TCACCTCAA CAACACCGCC ACCTTCCGGC TGCCCGCGA AACCGAGTTG  
 1201 CAAACCACTT TGGGCTTCAA TTATTTCCAC AACGAATACG GCAAAAACCG  
 1251 TTTCCTGAA GAATGGGGC TGTTCCTGA CGGTCTGAT CAGGACAACG  
 1301 GGCTTTATTC CTATTTGGGG CGGTTTAAGG GCGATAAAGG GCTGCTGCCC  
 1351 CAAAAATCAA CCATTGTCCA ACCGGCCGGC AGCCAATAT TCAACACGTT  
 1401 CTACTTCGAT GCCGCGCTCA AAAAAGACAT TTACCGCTTA AACTACAGCA  
 1451 CCAATACCGT CGGTACCGT TTCCGGCGCG AATATACGGG CTATTACGGC  
 1501 TCGGATGACG AATTTAAGCG GGCATTCGGA GAAAACTCGC CGACATACAA  
 1551 GAAACATTGC AACCGGAGCT GCGGGATTTA TGAACCCGTA TTGAAAAAT  
 1601 ACCGCAAAA GCGCGCCAAC AACCATTCCG TCAGCATTAG TCGGGAATTC  
 1651 GCGGATTAT TCATGCCGTT CGCCAGCTAT TCGGCGACAC ACCGTATGCC  
 1701 CAACATCCAA GAAATGTATT TTTCCCAAAT CCGCGACTCC GCGGTTTACA  
 1751 CCGCCTTAAA ACCAGAGCGC GCAAAACTT GGCATTTTGG CTTCATACC  
 1801 TATAAAAAAG GATTGTTAAA ACAAGATGAT ACATTAGGAT TAAACTGGT  
 1851 CGGCTACCGC AGCGCATCG ACACTACAT CCACAACGTT TACGGGAAT  
 1901 GGTGGGATTT GAACGGGGAT ATTCCGAGCT GGGTCAGCAG CACCGGGCTT  
 1951 GCCTACACCA TCCAACATCG CAATTTCAA GACAAAGTGC ACAAACACGG  
 2001 TTTTGAGTTG GAGCTGAATT ACGATTATGG CGGTTTTTTC ACCAACCTTT  
 2051 CTTACGCCA TCAAAAAAGC ACGCAACCGA CCAACTTCAG CGATGCGAGC  
 2101 GAATCGCCCA ACAATGCGTC CAAAGAAGAC CAACTCAAAC AAGGTTATGG  
 2151 GTTGAGCAGG GTTCCGCGCC TGCCGCGAGA TTACGGACGT TTGGAAGTCG  
 2201 TTACGCGCTG GTTGGGCAAC AAAGTACTT TGGGCGGCGC GATGCGCTAT  
 2251 TTCGGCAAGA GCATCCGCGC GACGGCTGAA GAACGCTATA TCACGCGCAC  
 2301 CAACGGGGGA AATACCAGCA ATTTCCGGCA ACTGGGCAAG CGTCCATCA  
 2351 ACAAACCGA AACTCTTGCC CGCCAGCCTT TGATTTTGA TTTTACGCG  
 2401 GCTTACGAGC CGAAGAAAA CCTTATTTTC CGCGCCGAAG TCAAAATCT  
 2451 GTTCGACAGG CGTTATATCG ATCCGCTCGA TGCGGGCAAT GATGCGGCAA  
 2501 CGCAGCGTTA TTACAGCTCG TTCGACCCGA AAGACAAGGA CGAAGACGTA  
 2551 ACGTGTAATG CTGATAAAAC GTTGTGCAAC GGCAAATACG GCGGCACAAG  
 2601 CAAAAGCGTA TTGACCAATT TTGACGCGG ACGCACCTTT TTGATGACGA  
 2651 TGAGCTACAA GTTTTAA

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50 1 EAQIQVLEDV HVKAKRVPKD KKVFTDARAV STRQDIFKSS ENLDNIVRSI  
 51 PGAFTQDDKS SGIVSLNIRG DSGFGRVNTM VDGITQTFYS TSTDAGRAGG  
 101 SSQFGASVDS NFIAGLDVVK GSFSGSAGIN SLAGSANLRT LGVDDVVQGN  
 151 NRYGLLLKGL TGTNSTKGNA MAAIGARKWL ESGASVGVLY GHSRRSVAQN  
 201 YRVGGGGQHI GNFGAEYLER RKQRYFVQEG ALKFNSDSGK WERDLQRQQW  
 55 251 KYKPYKNYNN QELQYIEEH DKSWRENLP QYDITPIDPS SLKQQSAGNL  
 301 FKLEYDGVFN KYTAQFRDLN TKIGSRKIIN RNYQFNYGLS LNPYTNLNT  
 351 AAYNSGRQKY PKGSKFTGWG LLKDFETYNN AKILDNLNTA TFRLPRETEL  
 401 QTTLGFNYFH NEYGKRFPE ELGLFFDGPD QDNGLYSYLG RFGDKGLLP  
 451 QKSTIVQPAG SQYFNTFYFD AALKKDIYRL NYSTNTVGYR FGGEYTGYYG  
 60 501 SDDEFKRAFG ENSPTYKKHC NRSCGIYEPV LKKYGKKRAN NHSVSIADF  
 551 GDYFMPFASY SRTHRPNIQ EMYFSQIGDS GVHTALKPER ANTWQGFNT  
 601 YKGLLLKQDD TLGLKLVGYR SRIDNYIHNV YGKWDLNGD IPSWVSSTGL  
 651 AYTIIHRNFK DKVHKHGFEL ELNYDYGRFF TNLSYAYQKS TQPTNFSAS  
 701 ESPNNASKED QLKQGYGLSR VSALPRDYGR LEVGTRWLGN KLTLLGAMRY  
 751 FGKSRATAE ERYIDGTNGG NTSNFRQLGK RSIKQTETLA RQPLIFDFYA  
 801 AYEPKKNLIF RAEVKNLFD RYIDPLDAGN DAATQRYYS FDKDKDEDV  
 851 TCNADKTLN GKYGTSKSV LTNFARGRTF LMTMSYKF\*

Computer analysis of this amino acid sequence gave the following results:



Homology with the probable TonB-dependent receptor HI121 of *H.influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

```

5      Orf133: 31  IYEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90
      HI121: 563  IEP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRMPNIQEM+FSQ+ ++GV+TA
      Orf133: 91  LKPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV 150
      HI121: 623  LKPEQSDTYQLGFENTYKKGLFTQDDVLGVKLVGYRSFIKNYIHNVYGWW--RDGMPTWA 680
10     Orf133: 151 SSTGLAYTIQHRXFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDAESPNN 210
      HI121: 681  S G YTI H+ + V YD GRFF N+SYAYQ++ QPTN++DAS PNN
      Orf133: 211 ASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYID 270
      HI121: 741  ASQEDILKQGYGLSRVSMPLPKDYGRLEGTWFDQKLTGLAARYYGKSKRATIEEYIN 800
15     Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDP 330
      HI121: 801  G+ + R+ ++K+TE + +QP+I D + +YEP K+LI +AEV+NL D+RY+DP
      Orf133: 331 LDAGNDAAXERYYSFDPKDKDXDVTNADKTLNKGKYGSTKSVLTNFARGRTFLMTMS 390
      HI121: 860  LDAGNDAASQRYSSL-----NNSIECAQDSSAC----GGSDKTVLYNFARGRTYILSLN 910
20     Orf133: 391 YKF 393
      HI121: 911 YKF 913
25     Orf133: 391 YKF 393
      HI121: 911 YKF 913
30     Orf133: 391 YKF 393
      HI121: 911 YKF 913

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N.*

*meningitidis*:

```

35     orf133.pep          10      20      30
      PGYYGSDDEFKRAFGENSPTXKKHKNRSCGI
      orf133a          FYFDAALKKDIYRLNYSTNTVGYRFGGXGTGYXSDDEFKRAFGENSPTYXKHKNQSCGI
      450      460      470      480      490      500
40     orf133.pep          40      50      60      70      80      90
      YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
      orf133a          YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
      510      520      530      540      550      560
45     orf133.pep          100      110      120      130      140      150
      KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS
      orf133a          KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVS
      570      580      590      600      610      620
50     orf133.pep          160      170      180      190      200      210
      STGLAYTIQHRXFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDAESPNN
      orf133a          STGLAYTIQHRNFKDKVHKHGFELNLDYXRFFTNLSYAYQKSTQPTNFSDAESPNN
      630      640      650      660      670      680
55     orf133.pep          220      230      240      250      260      270
      SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDG
      orf133a          SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDX
      690      700      710      720      730      740
60     orf133.pep          280      290      300      310      320      330
      TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL
      orf133a          TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL
      750      760      770      780      790      800

```

		1	AAAGACAAAA	AAGTGTTTAC	CGATGCGCGT	GCCGTATCGA	CCCGTCAGGA
		51	TATATTCAAA	TCCANGGAAA	ACCTCGACAA	CATCGTACGC	ANCATCCCCG
20		101	GTGCGTTCAT	ACANCAANAT	AAAAGCTCGG	GCNNTGTGTC	TTTGAATATT
		151	CGCNGCGACA	CGGGGTTCGG	GCGGGTCAAT	ACNATGTGTG	ACGGCATCAC
		201	NCANACCTTT	TATTCGACTT	CTACCGATGC	GGGCAGGGCA	GGCGGTTTCAT
		251	CTCAATTTCG	TGCATCTGTC	GACAGCAATT	TTATNGCCGG	ACTGGATGTC
		301	GTCAAAGGCA	GCTTCAGCGG	CTCGGCAGGC	ATCAACAGCC	TTGCCGGTTTC
25		351	GGCGAATCTG	CGGACTTTAN	GCGTGGATGA	TGTCGTTTCA	GCGCAATANTA
		401	CNTACGGCCT	GCTGCTAAAA	GGTCTGACCG	GCACCAATTG	AACCAAAGGT
		451	AATGCCATGG	CGGCGATAGG	TGCGCGCAAA	TGGCTGGAAA	GCGGAGCATC
		501	TGTCGGTGTG	CTTTACGGGG	ACAGCAGGGC	CAGCGTGGCG	GAAAATTACC
		551	GCGTGGGCGG	CGGCGGGCAG	CACATCGGAA	ATTTTGGCGC	CGAATATCTG
30		601	GAACGACGCA	AGCAACGATA	TTTTGAGCAA	GAAGGCGGGT	TGAAATTCAA
		651	TTCCAACAGC	GGAAAAATGG	AGCGGGATTT	CCAAAAGTTC	TACTGGAAAA
		701	CCAAGTGTTA	TCAAAAAATC	GATGCCCCCT	AAGAATGCA	AAAATACATC
		751	GAAGGTCAAT	ATCAAAAGCT	GCGGGAAAC	CTGGCGCCGC	AAATACGACAT
		801	CACCCCATC	GATCCGTCCA	GCCTGAAGCN	GCAGTCCGCA	GGCAACCTGT
35		851	TTAAATTGGA	ATACACGGCG	GTATTCAATA	AATACACGGC	GCAATTTCGC
		901	GATTTAAACA	CCAAAATCGG	CAGCCGCAAA	ATCATCAACC	GCAATTATCA
		951	ATTCAAATTAC	GGTTTGTCTT	TGAACCCGTA	TACCAACCTC	AATCTGACCG
		1001	CAGCCTACAA	PTCGGGCAGG	CAGAAATATC	CGAAAGGGTC	GAAGTTTACA
40		1051	GGCTGGGGGG	TTTTNAAAGA	TTTTGAAACC	TACAACAACG	CAAAAAATCT
		1101	CGACCTCANC	AACACTCCCA	CCTTCCGGCT	GCCCCGTGAA	ACCGAGTTGC
		1151	AAACCACTTT	GGGCTTCAAT	TATTTCCACA	ACGAATACGG	CAAAAACCGC
		1201	TTTCCTGAAG	AATTGGGGCT	GTTTTTCGAC	GGTCCGGATC	ANGACAACGG
		1251	GCTTTATTCC	TATTTGGGGC	GGTTTAAAGG	CGATAAAGG	CTGCTGCCCC
45		1301	AAAAATCAAC	CATTGTCCAA	CCGGCCGGCA	GCCAATTATT	CACACAGTTC
		1351	TACTTCGATG	CCGCGCTCAA	AAAAGACATT	TACCGCTTAA	ACTACAGCAC
		1401	CAATACCGTC	GGCTACCGTT	TCGGCGGCNA	ATATACGGGC	TATTACNGCT
		1451	CGGATGACGA	ATTTAAGCGG	GCATTCGGAG	AAAACCTGCC	GACATACANG
		1501	AAACATTGCA	ACCAAGAGCT	CGGAATTTAT	GAACCCGTAT	TGAAAAAATA
50		1551	CGGCAAAAAG	CGCGCCAACA	ACCATTCCGT	CAGCATTAGT	GCGGACTTCG
		1601	GCGATTATTT	CATCGCGTTC	GCCAGCTATT	CGCGCACACA	CCGTATGCCC
		1651	AACATCCAAG	AAATGTTATT	TTCCCAATC	GGCGACTCCG	GCGTTCCAC
		1701	CGCCTTAAAA	CCAGAGCGCG	CAAAACACTT	GCAATTTGCG	TTCAATACCT
		1751	ATAAAAAAGG	ATTGTTAAAA	CAAGATGATA	TATTAGGATT	AAAACCTGGT
55		1801	GGCTACCCGA	GCCGCATCGA	CNACTACATC	CACAACGTTT	ACGGGAAATG
		1851	GTGGGATTTG	AACGGGAATA	TTCCGAGCTG	GGTCAAGCAG	ACCGGGCTTG
		1901	CCTACACCAT	CCAACACCGC	AATTTCAAAG	ACAAAGTGCA	CAAACACGGT
		1951	TTTGAGTTGG	AGCTGAATTA	CGATTATNGG	CGTTTTTTCA	CCAACCTTTC
		2001	TTACGCCCTAT	CAAAAAAGCA	CGCAACCGAC	CAACTTCAGC	GATGCGAGCG
60		2051	AATCGGCCAA	CAATCGCTCC	AAAGAAGACC	AACTCAACA	AGGTTATGGG
		2101	TTGAGCAGGG	TTTCCGCCCT	GCCGCGAGAT	TACGGACGTT	TGGAAGTCGG
		2151	TACGCGCTGG	TTGGGGCAACA	AACTGACTTT	GGGCGGCGCG	ATGCGCTATT
		2201	TCGGCAAGAG	CATCCGCGCG	ACGGCTGAAG	AACGCTATAT	CGACGNCACC
		2251	AATGGGGNAN	TACCAGCAA	TTTCCGGCAA	CTGGGCAAGC	GTTCATCAN
65		2301	ACAAACCGAA	ACCCTTGCCC	GCCAGCCTTT	GATTTTTTGT	TTNTACGCCG
		2351	CTTACGAGCC	GAAGAAAAAN	CTTATTTTCC	GCGCCGAAGT	CAAAAATCTG
		2401	TTCGACAGGC	GTTATATCGA	TCCGCTCGAT	GCGGGCAATG	ATGCGGTCAT
		2451	GCAGCGTTAT	TACAGTTCGT	TCGACCCGAA	AGACAAGGAC	GAAGAAGTAA
		2501	CGTGTAATGA	TGATAACACG	TTATGCAACG	GCAAATACGG	GCGCACAGC
		2551	AAAAGCGTAT	TGACCAATTT	TGCACGCGGA	CNCACCTTTT	TGATAACGAT
		2601	GAGCTACAAG	TTTTTAA			

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

      1  KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI
      51  RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV
    101  VKGSFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG
    151  NAMAAGARK  WLESGASVGV LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL
    201  ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTQWYQKY DAPQELQKYI
    251  EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR
    301  DLNTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT
    351  GWGLXKDFET YNNAKILDIX NTSTFRLPRE TELQTLGFN YFHNEYGKNR
    401  FPEELGLFFD GPDNDGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF
    451  YFDAALKKDI YRLNYSTNTV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX
    501  KHCNQSCGIY EPVLKKYGKK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP
    551  NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLV
    601  GYRSRIDXYI HNVYQKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
    651  FELELNIDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGYG
    701  LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT
    751  NGXXTSNFRQ LGKRSIXQTE TLAQPLIFD XYAAEYEPKX LIFRAEVKNL
    801  FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNT LCNGKYGGTS
    851  KSVLTNFARG XTFLITMSYK F*
  
```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

                                     10      20      30      40
orfl33a.pep                        KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS
                                     |||||
orfl33-1      EAQIQVLEDVHVAKRVPKDKKVFTDARAVSTRQDIFKSSENLDNIVRSIPGAFTQQDKS
                                     10      20      30      40      50      60
                                     |||||
orfl33a.pep                        SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGSSQFGASVDSNFXAGLDVVK
                                     |||||
orfl33-1      SGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGSSQFGASVDSNFIAGLDVVK
                                     70      80      90      100     110     120
                                     |||||
orfl33a.pep                        GSFGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNAMAAGARKWL
                                     |||||
orfl33-1      GSFGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKGNAMAAGARKWL
                                     130     140     150     160     170     180
                                     |||||
orfl33a.pep                        ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
                                     |||||
orfl33-1      ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSGK
                                     190     200     210     220     230     240
                                     |||||
orfl33a.pep                        WERDFQKSYWKTQWYQKYDAPQELQKYEIEGHDKSWRENLAPQYDITPIDPSSLKXQSAGN
                                     |||||
orfl33-1      WERDLQRQWQKYPYKNYNN-QELQKYEIEGHDKSWRENLPQYDITPIDPSSLKQQSAGN
                                     250     260     270     280     290
                                     |||||
orfl33a.pep                        LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
                                     |||||
orfl33-1      LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
                                     300     310     320     330     340     350
                                     |||||
orfl33a.pep                        YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTLGFNYFHNEYGKNRFP
                                     |||||
orfl33-1      YPKGSKFTGWGLLKDFETYNNAKILDINNTATFRLPRETELQTLGFNYFHNEYGKNRFP
                                     360     370     380     390     400     410
                                     |||||
orfl33a.pep                        EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFFYFDAALKKDIYR
                                     |||||
orfl33-1      EELGLFFDGPQDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFFYFDAALKKDIYR
                                     420     430     440     450     460     470
  
```

-478-

		470	480	490	500	510	520
	orf133a.pep	LNYSTNTVGYRFGGXYTGYYXSDDEFKRAFGENSPITYKKHCNQSCGIYEPVLKKYGKKRA					
	orf133-1	LNYSTNTVGYRFGGEYTGYYGSDDEFKRAFGENSPITYKKHCNRSCGIYEPVLKKYGKKRA					
5		480	490	500	510	520	530
	orf133a.pep	NNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN					
10	orf133-1	NNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN					
		540	550	560	570	580	590
	orf133a.pep	TYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVSSSTGLAYTIQHRNF					
15	orf133-1	TYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVSSSTGLAYTIQHRNF					
		600	610	620	630	640	650
	orf133a.pep	KDKVHKHGFLELNYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS					
20	orf133-1	KDKVHKHGFLELNYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS					
		660	670	680	690	700	710
	orf133a.pep	RVSALPRDYGRLVGTRWLGKLTGGAMRYFGKSIRATAEERYIDXTNGXXTSNFRQLG					
25	orf133-1	RVSALPRDYGRLVGTRWLGKLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNFRQLG					
		720	730	740	750	760	770
	orf133a.pep	KRSIXQTETLARQPLIFDYAAYEPKKKLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS					
30	orf133-1	KRSIXQTETLARQPLIFDYAAYEPKKKLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS					
		780	790	800	810	820	830
35	orf133a.pep	SFDPKDKDEEVTCTNDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX					
	orf133-1	SFDPKDKDEEVTCTNDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX					
40		840	850	860	870	880	

Homology with a predicted ORF from *N.gonorrhoeae*ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*45 *gonorrhoeae*:

	orf133.pep	PGYYGSDDEFKRAFGENSPITYKKHCNRSCGI	31
	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL	560
50	orf133.pep	YEPVLKKYGKKRANNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYGKKRANNHSVSISADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
55	orf133.pep	KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
60	orf133.pep	STGLAYTIQHRXFXDKVHQXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	211
	orf133ng	STGLAYTIRHRNFKDKVHKHGFLELNYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLVGTRWLGKLTGGAMRYFGKSIRATAEERYIDG	271
65	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLVGTRWLGKLTGGAMRYFGKSIRATAEERYIDG	800
	orf133.pep	TNGGNTSNFRQLGKRSIXQTETLARQPLIXDFNAAYEPKKKLI FRAEVKNLFDRRYIDPL	331
	orf133ng	TNGGNTSNVRQLGKRSIXQTETLARQPLIFDYAAYEPKKKLI FRAEVKNLFDRRYIDPL	860

```

orf133.pep      DAGNDAAXERYSSFDPKDKDXDVTCNADKTLGNGKYGGTSKSVLTNFARGRTFLMTMSY  391
                |||||::|||
orf133ng        DAGNDAATQRYSSFDPKDKDEDVTCNADKTLGNGKYGGTSKSVLTNFARGRTFLMTMSY  920
orf133.pep      KF  393
                ||
orf133ng        KF  922

```

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a protein having amino acid sequence <SEQ ID 882>:

15

20

25

30 A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

35

40

45

50

55

60

65

2.5

30

35

40

45

50

55

60

65

1	MRSSFRLKPI	CFYLMGVMLY	HHSYAEDAGR	AGSEAIQIVL	EDVHVKAQRV
51	PKDKKVFTDA	RAVSTRQDVF	KSGENLDNIV	RSIPGAFTQO	DKSSSGIVSLN
101	IRGDSGFGVR	NTMVDGITQT	FYSTSTDAGR	AGGSSQFGAS	VDSNFIAGLD
151	VVKGSFSGSA	GINSLAGSAN	LRTLGVDDVV	QGNNTYGLLL	KGLTGTNSTK
201	GNAMAAIGAR	KWLESGASVG	VLYGHSRRGV	AQNYRVGGGG	QHIGNFGEEY
251	LERRQQQYFV	QEGGLKFNAG	SGKWERDLQR	QYWKTKWYKK	YEDPQELQKY
301	IEEHDKSWRE	NLAPQYDITP	IDPSGLKQOS	AGNLFKLEYD	GVFNKYTAQY
351	RDLNLTRIGSR	KIINRNYQFN	YGLSLNPYTN	LNLTAAYNYS	RQKYPKGAKF
401	TGWGLLKDFE	TYNNAKILDL	NNTATFRLPR	ETELQTTLGF	NYFHNEYGKN
451	RFPEELGLFF	DGPDQDNGLY	SYLGRFGDGK	GLLPQKSTIV	QPAGSQYFNT
501	FYFDAAALKD	IYRLNYSYTA	INYRFGEYTT	GYGSENEFEK	RAFGENSPAY
551	KEHCDPSCGL	YEPVLKYGKG	KRANNHVSIS	SADFGDYFMP	FAGYSRTHRM
601	PNIQEMYFSQ	IGDSVGHTAL	KPERANTWQF	GFNTYKKGLL	KQDDILGLKL
651	VGYRSRIDNY	IHNVYGKWWD	LNGDIPSWVG	STGLAYTIRH	RNFKDKVHKH
701	GFELELNNDY	GRFFTNLSYA	YQKSTQPTNF	SDASESPNNA	SKEQDLQKGY
751	GLSRVSALPR	DYGRLEVGTG	WLGNKLTLLG	AMRYFGKSIR	ATAEERYIDG
801	TNGGNTSNVR	DLGKRSIKQT	ETLARAQPLIF	DFYAAYPEKK	NLIFRAEVKN
851	LKDRRYIDPL	DAGNDAATQR	YSSYDFDPDK	DEDVTCNADK	TLCNGKYGGT
901	SFSVLTNTEF	GRTFLMTMSY	KF*		

	10	20	30	40	50	60							
orf133ng-1.pep	SFRLK	PICFYLMGV	MLYHH	SHYAEDAG	RAGSEAQ	IQVLE	DVHV	KA	KRV	PKDK	KVFT	DARAV	
orf133-1													
					EAQIQ	VLE	DVHV	KA	KRV	PKDK	KVFT	DARAV	
						10			20			30	
	70	80	90	100	110	120							
orf133ng-1.pep	STRQ	DVPFK	SGENL	DNIVR	SIPGA	FTQQD	KSSGI	VSLN	IRGDS	SGFGR	VNTM	VDGIT	QTFFYS
orf133-1		:	:										
	STRQ	DIFK	SSENL	DNIVR	SIPGA	FTQQD	KSSGI	VSLN	IRGDS	SGFGR	VNTM	VDGIT	QTFFYS
		40		50		60		70		80		90	
	130	140	150	160	170	180							
orf133ng-1.pep	TSTD	AGRAGG	SSQFG	ASVDS	NFIAG	LDDV	VKGS	FSGS	SAGIN	SLAGS	ANLRT	LGVD	DDVQGN
orf133-1													
	TSTD	AGRAGG	SSQFG	ASVDS	NFIAG	LDDV	VKGS	FSGS	SAGIN	SLAGS	ANLRT	LGVD	DDVQGN
		100		110		120		130		140		150	
	190	200	210	220	230	240							
orf133ng-1.pep	NTYGL	LLKGL	TGTN	STKGN	AMAAI	GARKW	LES	SGASV	GVLY	GHSRR	GAQ	NYRV	GGGGQHI
orf133-1													
	NTYGL	LLKGL	TGTN	STKGN	AMAAI	GARKW	LES	SGASV	GVLY	GHSRR	GAQ	NYRV	GGGGQHI
		160		170		180		190		200		210	
	250	260	270	280	290	300							
orf133ng-1.pep	GNFG	EYLER	RKQQY	FVQEG	GLKFN	AGSGK	WERDL	QRQY	WKT	KWKY	KYEDP	QELQ	KYIEE

-481-

```

||||| |||||:|||||: ||||| ||| | :||: |||||
orf133-1  GNFGAEYLERRKQRYFVQEGALKFNSDSGKWERDLQRQQWKYPYKNNYNN-QELQKYIEE
          220      230      240      250      260

5      310      320      330      340      350      360
orf133ng-1.pep HDKSWRENLAPOYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
          ||||| |||||:|||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      HDKSWRENLAPOYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
          270      280      290      300      310      320

10     370      380      390      400      410      420
orf133ng-1.pep NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKSKFTGWGLLKDFETYNNAKILDNLNT
          330      340      350      360      370      380

15     430      440      450      460      470      480
orf133ng-1.pep ATFRLEPRETELQTTLGFNYPFHNEYGKNRFPPEELGLFFDGPDQDNGLYSYLGRFKGDKGLL
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      ATFRLEPRETELQTTLGFNYPFHNEYGKNRFPPEELGLFFDGPDQDNGLYSYLGRFKGDKGLL
          390      400      410      420      430      440

20     490      500      510      520      530      540
orf133ng-1.pep PQKSTIVQPAGSQYFNTFFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAF
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      PQKSTIVQPAGSQYFNTFFYFDAALKKDIYRLNYSTNTVGYRFGGEYTGYYGSDDEFKRAF
          450      460      470      480      490      500

25     550      560      570      580      590      600
orf133ng-1.pep GENSPAYKEHCDPSCGLYEPVLKKYGGKRRNNHVSISADFGDYFMPFAGYSRTHRMPNI
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      GENSPYKHKCNRSCTIYEPVLKKYGGKRRNNHVSISADFGDYFMPFASYSRTHRMPNI
          510      520      530      540      550      560

30     610      620      630      640      650      660
orf133ng-1.pep QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDILGLKLVGYRSRIDNYIHN
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLVGYRSRIDNYIHN
          570      580      590      600      610      620

35     670      680      690      700      710      720
orf133ng-1.pep VYGKWWDLNGDIPSWVGSTGLAYTIHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      VYGKWWDLNGDIPSWVSSSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
          630      640      650      660      670      680

40     730      740      750      760      770      780
orf133ng-1.pep STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWRLGNKLTGGAMR
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWRLGNKLTGGAMR
          690      700      710      720      730      740

45     790      800      810      820      830      840
orf133ng-1.pep YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAEYEPKKNLI
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      YFGKSIRATAEERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLIFDYAAEYEPKKNLI
          750      760      770      780      790      800

50     850      860      870      880      890      900
orf133ng-1.pep FRAEVKNLFDRLYIDPLDAGNDAATQRYSSFPDKDKDEDVTCNADKTLNCGKYGGTSKS
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      FRAEVKNLFDRLYIDPLDAGNDAATQRYSSFPDKDKDEDVTCNADKTLNCGKYGGTSKS
          810      820      830      840      850      860

55     910      920
orf133ng-1.pep VLTNFARGRTFLMTMSYKFX
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf133-1      VLTNFARGRTFLMTMSYKFX
          870      880

```

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

-482-

sp|P45114|YC17\_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR H11217 PRECURSOR  
>gi|1075372|pir|G64110 transferrin binding protein 1 precursor (tbpl) homolog -  
Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding  
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913  
Score = 930 bits (2377), Expect = 0.0  
Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

5  
Query: 38 QVLEDVHVKA KRVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97  
+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V  
10  
Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKSGSVV 88

Query: 98 SLNIRGDSGFGVRVNTMVDGITQTFYSTSTDA GRAGSSQFGASVDSNFIAGLDVVKGSFS 157  
S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GSSQFGA++D NFIAG+DV K +FS  
15  
Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGGSSQFGAIDPNFIAGVDVNKSNFS 148

Query: 158 GSAGINSLAGSANLRTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217  
G++GIN+LAGSAN RTLGV+DV+ M RKWL++G  
Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRKWLDNGG 208

20  
Query: 218 SVGVLYGHSRRGVAQNYRVGGGQHGNEEGEYLERRKQYFVQEGGLKFNAGSGKWERD 277  
VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D  
Sbjct: 209 YGVVYGYVSQREVSDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYIILNP-EGQWTPD 265

25  
Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303  
L +++W +Y KK +D ++LQK IEE  
Sbjct: 266 LSKKHWSCNKPDPYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDIEE 325

30  
Query: 304 HDKSWRENLAPOQYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363  
DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI  
Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQNGLGAQLRTLONKIGSRKIE 384

35  
Query: 364 NRNYQFN YGLSLNPTYNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDLNNT 423  
NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+  
Sbjct: 385 NRNYQVNYNFNNNSYLDNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444

40  
Query: 424 ATFRLPRETELQTTLGFN YFHNEYGKNRFPPEELGLFFDGPDDNGLYSY--LGRFKGDKG 481  
TF LP+E +L+TTLGFNYF NEY KNRFPPEEL LF++ D GLYS+ GR+ G K  
Sbjct: 445 HTFLLPKEIDLKTTLGFN YFTNEYSKNRFPPEELSLFYNDASHDQGLYSHSKRGYSGTGS 504

45  
Query: 482 LLPQKSTIVQPAGSQYFNTFYFDAALKKDIYRLNYS TNAINYRFGGEYTGYYGSENEFKR 541  
LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY  
Sbjct: 505 LLPQRSVILQPSGKQKFTVYFD TALSGIYHLNYSVNFTHYAFNGEYVGY----- 555

50  
Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKGKRRANNHVSISADFGDYFMPFAGYSRTHRMP 601  
EN+ + + EP+L K G K+A NHS ++SA+ DYEMPF YSRTHRMP  
Sbjct: 556 ---ENTAGQQ-----INEPILHKS GHKKA FNHSATLSAELSDYFMPFFTYSRTHRMP 604

55  
Query: 602 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661  
NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKKGL QDD+LG+KLVGYRS I NYI  
Sbjct: 605 NIQEMFFSQVSNAGVN TALKEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNIYI 664

60  
Query: 662 HN VY GKW WDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNYDYGRFFTNLSYAY 721  
HN VY G WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY  
Sbjct: 665 HN VY GVW W--RDGMPTWAESNGFKYTI AHQNYKPIVKKSGVELEINYDMGRFFANVSYAY 722

65  
Query: 722 QKSTQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLGNKLT LGGA 781  
Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLT LG A  
Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMLPKDYGRLELGT RWFDQKLT LGLA 782

70  
Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAA YEPKKN 841  
RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+  
Sbjct: 783 ARYYGKSKRATIEEYINGSR-FKKNLTLRENYAYVKKTEDIKKQPIILD LHVSYEPIKD 841

Query: 842 LIFRAEVKNLFDRRYIDPLDAGNDAATQRYYS SFDPKDKDEDVTCNADKTL CNKG YGGS 901  
LI +AEV+NL D+RY+DPLDAGNDAA+QRYYS + + C D + C GG+  
Sbjct: 842 LI IKA EVQNLLDKRYVDPLDAGNDAASQRYYS L-----NNSIECAQDSSAC----GGSD 892

Query: 902 KSVLTNFARGRTFLMTMSYKF 922  
K+VL NFARGRT+++++YKF  
Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913



The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1  ATGAACCTGA TTTACAGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51  TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151 GGCTACACCG CCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATCCCCCT
201 CGCCGTCTTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAG AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
351 CCGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1  MNLISRYIIR QMAVMAYVAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51  GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTL SQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSVINVR EMLPDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1  ATGAACCTGA TTTACAGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51  TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151 gGCTACACCG CCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATCCCCCT
201 CGCCGTCTTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAG AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
351 CCGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCTTKAT CAATGTGCGC GAAATGTTGC CCGACCATAC
501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAC GAATTGGCAG
551 AGGCAGTGGA AGCCGATTCC GCCGTTTTGA ACAGCGACGG CAGTTGGCAG
601 TTGAAAAACA TCCGCCGCAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC
651 TATTGCGGCT GAAGAAAAC TGGCCGATTTC CGTCAAACGC AACCTGATGG
701 ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC
751 TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCCGAA TCTACGCCAT
801 CGCATGGTGG CGCAAATTGG TTTACCCCGC CGCAGCCTGG GTGATGGCGC
851 TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC
901 TTAATACTCT TCGGCGGCAT CTGTsTCGGA TTGCTGTTCC ACCTTGCCCG
951 ACGGCTCTTT GGGTTTACCA GCCAACTCGG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1  MNLISRYIIR QMAVMAYVAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51  GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTL SQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSXINVR EMLPDHLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
201 LKNIRRLSTLG EDKVEVSIAA EENWPISVVR NLMDVLLVKP DQMSVGELTT
251 YIRHLQNSQ NTRIYAIAWW RKLVPYAAAW VMLVAFAFT PQTTTRHGNMG
301 LKLFGGICXG LFLHLAGRLF GFTSQL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

-484-

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N. meningitidis*:

5	orf112.pep	10 20 30 40 50 60	MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
	orf112a	10 20 30 40 50 60	MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
10	orf112.pep	70 80 90 100 110 120	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLSQFGFIFAIATVALGEW
	orf112a	70 80 90 100 110 120	AYELMPLAVLIGGLVXSQAAGSELXVIKASGMSTKKLLLSQFGFIFAIATVALGEW
15	orf112.pep	130 140 150 160	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH
	orf112a	130 140 150 160	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
20	orf112a	190 200 210 220 230 240	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKNRLMDVLLVKP

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
30	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCGC	GCCTACGAAC	TGATGCCCCT
35	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	GCCGCCGGCA
	251	GCGAACTGAN	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
40	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTTT	GCTATTGCCA	CCGTCGCGCT
	351	CGGCGAATGG	GTTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
45	401	CCGCGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
50	501	CCTGCTGGGC	ATTAAATCT	GGGCCCGCAA	CGATAAAAC	GAAGTGGCAG
	551	AGGCAGTGGA	AGCCGATTCC	GCCGTTTGA	ACAGCGACGG	CAGTTGGCAG
55	601	TTGAAAAACA	TCCGCCGCAG	CACGCTTGGC	GAAGACAAAG	TGAGGTCTC
	651	TATTGCGGCT	GAAGAAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
60	701	ACGTATTGCT	CGTCAAACCC	GACCAAATGT	CCGTGCGCGA	ACTGACCACC
	751	TACATCCGCC	ACCTCCAAN	NNACAGCCAA	AACACCCGAA	TCTACGCCAT
65	801	CGCATGGTGG	CGCAAATTGG	TTACCCCGC	CGCAGCCTGG	GTGATGCGCG
	851	TCGTGCGCTT	TGCCTTTACC	CCGCAAACCA	CCCGCCACGG	CAATATGGGC
70	901	TTAAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTC	ACCTTGCCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGGCATCCCG	CCCTTCCTCG
75	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

50	1	MNLISRYIIR	QMAVMNAVYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX
	51	GYTALKMXAR	AYELMPLAVL	IGGLVXSQ	AAGSELXVIK	ASGMSTKKLL
55	101	LILSQFGFIF	AIATVALGEW	VAPTLSQKAE	NIKAAAINGK	ISTGNTGLWL
	151	KEKNSIINVR	EMLPDHTLLG	IKIWARNDKN	ELAEAVEADS	AVLNSDGSWQ
60	201	LKNIRRSTLG	EDKVEVSIAA	EEXWPISVVR	NLMDVLLVKP	DQMSVGELTT
	251	YIRHLQXXSQ	NTRIYAIWW	RKLVPAAAW	VMALVAFAT	PQTRHGNMG
65	301	LKXFGGICLG	LLFHLAARLF	XFTSQLYGIP	PFLXGALPTI	AFALLAVWLI
	351	RKQEK*				

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

60	orf112a.pep	MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
	orf112-1	MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
65	orf112a.pep	AYELMPLAVLIGGLVXSQAAGSELXVIKASGMSTKKLLLSQFGFIFAIATVALGEW

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N. gonorrhoeae*:

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

This encodes a protein having amino acid sequence <SEQ ID 892>:

60

1	<u>MNLISRYIIR</u>	<u>QMAVMVAYAL</u>	<u>LAFLALYSFF</u>	<u>EILYETGNLG</u>	<u>KGSYGIWEML</u>
51	<u>GYTALKMPAR</u>	<u>AYELMPLAIVL</u>	<u>IGGLASLSQL</u>	<u>AAGSELAIVK</u>	<u>ASGMSTPKLL</u>
101	<u>LILSQFGFIF</u>	<u>ALAAVALGEW</u>	<u>VAPTLTSQAKE</u>	<u>NIKAAAINKG</u>	<u>ISTGNTGLWL</u>
151	<u>KEKTSINVR</u>	<u>GMLPDLTLLG</u>	<u>IKIWARNDKN</u>	<u>ELAAEVEADS</u>	<u>AVLNSDGSQW</u>
201	<u>LKNIRRSIMG</u>	<u>TDKIETSAAA</u>	<u>EETWPIAVRR</u>	<u>NLMDVLLVKP</u>	<u>DQMSYGEHLT</u>
251	<u>YIRHLQNSNG</u>	<u>NTQIYAIAWW</u>	<u>RKLVPYVAAW</u>	<u>VMLVAFAFT</u>	<u>PQTRRHGNGM</u>
301	<u>LKLFEGGICLG</u>	<u>LFLHLAGRLF</u>	<u>GFTSOLYGTP</u>	<u>PFALAGLPTI</u>	<u>AFALLAWVLI</u>

-486-

351 RKQEK R\*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLISRYIIRQMAVM	AVYALLAFLALYSF	FEILYETGNLGKGS	YGIWEMLGYTALK	MPAR	
	orf112-1	MNLISRYIIRQMAVM	AVYALLAFLALYSF	FEILYETGNLGKGS	YGIWEMLGYTALK	MPAR	
		10	20	30	40	50	60
10	orf112ng	AYELMPLAVLIGGL	ASLSQLAAGSELAV	IKASGMSTKKLL	LILSQFGFIFAIA	AAVALGEW	
	orf112-1	AYELIPLAVLIGGL	VLSLSQLAAGSEL	TVIKASGMSTKK	LLILSQFGFIFA	IATVALGEW	
		70	80	90	100	110	120
15	orf112ng	VAPTLSQKAENIKAA	AINGKISTGNTGL	WLKEKTSIINVR	GMLPDHTLLGI	KIWARNDKN	
	orf112-1	VAPTLSQKAENIKAA	AINGKISTGNTGL	WLKEKTSIINVR	GMLPDHTLLGI	KIWARNDKN	
		130	140	150	160	170	180
20	orf112ng	ELAEAVEADSAVL	NSDGSWQLKNIR	RSIMGTDKIETSA	AAEETWPIAVR	RNLMDVLLVKP	
	orf112-1	ELAEAVEADSAVL	NSDGSWQLKNIR	RSTLGEDKVEVSI	AAEENWPISVK	RNLMDVLLVKP	
		190	200	210	220	230	240
25	orf112ng	DQMSVGELTTYIR	HLQNNSQNTQIYA	IAWWRKLVYPVA	AAWVMALVAF	AFTPQTRHGNMG	
	orf112-1	DQMSVGELTTYIR	HLQNNSQNTQIYA	IAWWRKLVYPAA	AWVMALVAF	AFTPQTRHGNMG	
		250	260	270	280	290	300
30	orf112ng	LKLFGGICLGLF	HLAGRLFGFTSQ	LYGTPPFLAGAL	PTIAFALLAV	WLIRKQEKRX	
	orf112-1	LKLFGGICLGLF	HLAGRLFGFTSQ	LYGTPPFLAGAL	PTIAFALLAV	WLIRKQEKRX	
		310	320	330	340	350	
35	orf112ng	LKLFGGICLGLF	HLAGRLFGFTSQ	LYGTPPFLAGAL	PTIAFALLAV	WLIRKQEKRX	
	orf112-1	LKLFGGICLGLF	HLAGRLFGFTSQ	LYGTPPFLAGAL	PTIAFALLAV	WLIRKQEKRX	
		310	320	330	340	350	

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTTCGG	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT	
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG	
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGATTTACGGACGCGCA	
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC	
ORF 5	Forward	GGAATTCATATGGCCATGG-TGGAAGGCGCACAAACC	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAAC	
	Reverse	CCCGCTCGAG-GACTGTGCAAAAACGG	
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC	
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTTGTTC	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTTGCGGA	
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGGCGC	
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTCGAGGTTTTCGGG	
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TAATGGGAAACCTTGTTTT	
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC	
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCGCGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG	
ORF 15	Forward	GGAATTCATATGGCCATGG-GCGGGACACTGACAG	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-TGCGGGACACTGACAGG	
	Reverse	CCCGCTCGAG-AGGTTGGCCTTGCTCTATG	
ORF 17	Forward	GGAATTCATATGGCCATGG -TTGCCGGCCTGTTTCG	NdeI-NcoI

	Forward	C <u>GGGATCC</u> -ATTGCCGGCTGTTCCG	BamHI
	Reverse	CCCGCTC <u>GAG</u> -AAGCAGTTGTACAGC	XhoI
<b>ORF 18</b>	Forward	GCGGATCCCATATG-ATTTGCTGCATTGGAT	BamHI-NdeI
	Reverse	CCCGCTC <u>GAG</u> -TCTTCCAATTTCTGAAAGC	XhoI
<b>ORF 19</b>	Forward	GGAATTCATATGGCCATGG -TCGCCAGTGT TTTTACC	NdeI-NcoI
	Forward	C <u>GGGATCC</u> -TTCGCCAGTGT TTTTACCG	BamHI
	Reverse	CCCGCTC <u>GAG</u> -GGTGT TTTTGAAGCTGCC	XhoI
<b>ORF 20</b>	Forward	GGAATTCATATGGCCATGG -TCGGCGCGGGTATG	NdeI-NcoI
	Forward	C <u>GGGATCC</u> -TTCGGCGCGGGTATG	BamHI
	Reverse	CCCGCTC <u>GAG</u> -CGGCGAGCGAGAGCA	XhoI
<b>ORF 22</b>	Forward	GGAATTCATATGGCCATGG-TGATTAAATCAAAAAAGGTCT	NdeI-NcoI
	Forward	C <u>GGGATCC</u> -ATGATTAAATCAAAAAAGGTCTAAACC	BamHI
	Reverse	CCCGCTC <u>GAG</u> -ATTATGATAGCGGCC	XhoI
<b>ORF 23</b>	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC	BamHI-NdeI
	Reverse	CCCGCTC <u>GAG</u> -TTTAAACCGATAGGTAAACG	XhoI
<b>ORF 24</b>	Forward	GGAATTCATATGGCCATGG -TGATGCCGGAATGGTG	NdeI-NcoI
	Forward	C <u>GGGATCC</u> -ATGATGCCGGAATGGTG	BamHI
	Reverse	CCCGCTC <u>GAG</u> -TGTCAGCGTGGCGCA	XhoI
<b>ORF 25</b>	Forward	GCGGATCCCATATG-TATCGCAAACGATTGC	BamHI-NdeI
	Reverse	CCCGCTC <u>GAG</u> -ATCGATGGAATAGCCG	XhoI
<b>ORF 26</b>	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC	BamHI-NdeI
	Reverse	CCCGCTC <u>GAG</u> -GACATCGGCGCGTTTT	XhoI
<b>ORF 27</b>	Forward	GGAATTCATATGGCCATGG-AGACCTATTCTGTTTA	NdeI-NcoI
	Forward	C <u>GGGATCC</u> - CAGACCTATTCTGTTTATTTAATC	BamHI
	Reverse	CCCGCTC <u>GAG</u> -GGGTTCGATTAAATAACCAT	XhoI
<b>ORF 28</b>	Forward	GGAATTCATATGGCCATGG-ACGGCTGTACGTTGATGT	NdeI-NcoI
	Forward	C <u>GGGATCC</u> -AACGGCTGTACGTTGATG	BamHI
	Reverse	CCCGCTC <u>GAG</u> -TTTGTCTAGAGGAATTCGCG	XhoI
<b>ORF 29</b>	Forward	GCGGATCCCATATG -AACGGTTTGGATGCCCCG	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-AACGGTTTGGATGCCCCG	BamHI-NheI
	Reverse	CCCGCTC <u>GAG</u> -TTTGTCTAAGTTCCTGATATG	XhoI
<b>ORF 32</b>	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTTG	BamHI-NdeI
	Reverse	CCCGCTC <u>GAG</u> -GCGTATTTTGTGCTTTG	XhoI
<b>ORF 33</b>	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG	BamHI-NdeI
	Reverse	CCCGCTC <u>GAG</u> -TTGATCTTTCAAACGGCC	XhoI

<b>ORF 35</b>	Forward Forward Reverse	GCGGATCCCATATG-TTCAGAGCTCAGCTT CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT CCCGCTCGAG-AAACAGCCATTTGAGCGA	BamHI-NdeI BamHI-NheI XhoI
<b>ORF 37</b>	Forward Reverse	GCGGATCCCATATG-GATGACGTATCGGATTTT CCCGCTCGAG-ATAGCCCGCTTTCAGG	BamHI-NdeI XhoI
<b>ORF 58</b>	Forward Reverse	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT CCCGCTCGAG-AGCATTGTCCAAGGGGAC	BamHI-NheI XhoI
<b>ORF 65</b>	Forward  Forward Reverse	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG  CGGGATCC-TTGCTGTATCTGAATCAAGG CCCGCTCGAG-CCGCATCGGCAGACA	  NdeI-NcoI BamHI XhoI
<b>ORF 66</b>	Forward Reverse	GCGGATCCCATATG-TACGCATTTACCGCCG CCCGCTCGAG-TGGATTTTGAGAGATGG	BamHI-NdeI XhoI
<b>ORF 72</b>	Forward Reverse	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA CCCGCTCGAG-GCCTGAGACCTTTGCAA	BamHI-NdeI XhoI
<b>ORF 73</b>	Forward Reverse	GCGGATCCCATATG-AGATTTTTCGGTATCGG CCCGCTCGAG-TTCATCTTTTTCATGTTG	BamHI-NdeI XhoI
<b>ORF 75</b>	Forward Reverse	GCGGATCCCATATG- TCTGTCTTTCAAACGGC CCCGCTCGAG-TTGTTTTTGCAAGACAG	BamHI-NdeI XhoI
<b>ORF 76</b>	Forward Reverse	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC CGGGATCC-TTACGGTTTGACACCGTT	NheI-NdeI BamHI
<b>ORF 79</b>	Forward Reverse	CGCGGATCCCATATG-GTTTCCGCCGCCG CCCGCTCGAG-GTGCTGATGCGCTTCG	BamHI-NdeI XhoI
<b>ORF 83</b>	Forward Reverse	GCGGATCCCATATG-AAAACCCTGCTGCTGC CCCGCTCGAG-GCCGCCTTTGCGGC	BamHI-NdeI XhoI
<b>ORF 84</b>	Forward Reverse	GCGGATCCCATATG-GCAGAGATCTGTTTG CCCGCTCGAG-GTTTGCCGATCCGACCA	BamHI-NdeI XhoI
<b>ORF 85</b>	Forward Reverse	CGCGGATCCCATATG- GCGGTTTGGGGCGGA CCCGCTCGAG-TCGGCGCGGCGGGC	BamHI-NdeI XhoI
<b>ORF 89</b>	Forward Forward Reverse	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA CGGGATCC-GCCATACCTTCTTATCAGAG CCCGCTCGAG-TTTTTCGATTAGAAAAAGC	NdeI-NcoI BamHI XhoI
<b>ORF 97</b>	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-NdeI

<b>ORF 98</b>	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
	Forward Reverse	GCGGATCCCATATG-ACGGTAACTGCGG CCCGCTCGAG-TTGTGTTCGGGCAAATC	BamHI-NdeI XhoI
<b>ORF 100</b>	Forward Reverse	GCGGATCCCATATG-TCGGGCATTACACCG CCCGCTCGAG-ACGGGTTTCGGCGGAA	BamHI-NdeI XhoI
<b>ORF 101</b>	Forward Reverse	GCGGATCCCATATG-ATTTATCAAAGAAACCTC CCCGCTCGAG-TTTCGCGCTTTCAATGT	BamHI-NdeI XhoI
<b>ORF 102</b>	Forward Reverse	GCGGATCCCATATG-GCAGGGCTGTTTTACC CCCGCTCGAG-AAACGGTTTGAACACGAC	BamHI-NdeI XhoI
<b>ORF 103</b>	Forward Reverse	GCGGATCCCATATG-AACCACGACATCAC CCCGCTCGAG-CAGCCACAGGACGGC	BamHI-NdeI XhoI
<b>ORF 104</b>	Forward Reverse	GCGGATCCCATATG-ACGTGGGGAACGC CCCGCTCGAG-GCGGCGTTTGAACGGC	BamHI-NdeI XhoI
<b>ORF 105</b>	Forward Reverse	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC CCCGCTCGAG-TAAACGAATGCCGTCCAG	BamHI-NdeI XhoI
<b>ORF 106</b>	Forward Reverse	GCGGATCCCATATG-AGGATAACCGACGGCG CCCGCTCGAG-TTGTTCCTCGATGATGT	BamHI-NdeI XhoI
<b>ORF 109</b>	Forward Reverse	GCGGATCCCATATG-GAAGATTTATATATAATACTCG CCCGCTCGAG-ATCAGCTTCGAACCGAAG	BamHI-NdeI XhoI
<b>ORF110</b>	Forward Reverse	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC AAACTGCAG-GGAAAACCACATCCGCACTCTGCC	EcoRI PstI
<b>ORF111</b>	Forward Reverse	AAAGAATTC-GCACCGCCAAAAGGCAAAAACCGCA AAACTGCAG-TCTGCGCGT TTTGCGGCAGGGTGG	EcoRI PstI
<b>ORF113</b>	Forward Reverse	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG AAACTGCAG-TTACGAATGCCTGCTTGTCTGACCGTACTG	EcoRI PstI
<b>ORF115</b>	Forward Reverse	AAAGAATTC-TTGCTTGTGCAACAGAAAAAGACGG AAAAAAGTCGAC-CTATTTTTTAGGGCTTTTGC TTGTTGAAAAGCCTGCC	EcoRI SalI
<b>ORF119</b>	Forward Reverse	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTTGCC	EcoRI PstI
<b>ORF120</b>	Forward Reverse	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG AAACTGCAG-CGGTTTTGGCTGCCTGGCCGTGAT	EcoRI PstI
<b>ORF121</b>	Forward Reverse	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC	EcoRI PstI



<b>ORF122</b>	Forward	AAAAAAGTCGAC-ATGTCCTACCGCGCAAGCAGTTCCTCC	SalI
	Reverse	AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC	PstI
<b>ORF125</b>	Forward	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT	EcoRI
	Reverse	AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG	PstI
<b>ORF126</b>	Forward	AAAGAATTC-GCGGAAACGGTCGAAG	EcoRI
	Reverse	AAACTGCAG-TTAATCTTGTCTCCGATATAC	PstI
<b>ORF127</b>	Forward	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG	EcoRI
	Reverse	AAAAAAGTCGAC-CTTAAGTAACTGCAGTCCTTATC	SalI
<b>ORF128</b>	Forward	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC	EcoRI
	Reverse	AAACTGCAG-CTATTGCAATGCGCCGCCGGAATGTTGAGCAGGCG	PstI
<b>ORF129</b>	Forward	AAAGAATTC-ATGGATTTTCGTTTTGACATTATTTACGAATACCG	EcoRI
	Reverse	AAACTGCAG-TTATTTTTTGATGAAATTTTGGGCGG	PstI
<b>ORF130</b>	Forward	AAAGAATTC-GCAGTACTTGCCAT TCTCGGTGCG	EcoRI
	Reverse	AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT	PstI
<b>ORF 131</b>	Forward	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT	BamHI-NdeI
	Reverse	CCCGCTCGAG-CCAGCGACGCGTTC	XhoI
<b>ORF 132</b>	Forward	GCGGATCCCATATG-AAAGAAGCGGGGTTTG	BamHI-NdeI
	Reverse	CCCGCTCGAG-CCAATCTGCCAGCCGT	XhoI
<b>ORF 133</b>	Forward	CGCGGATCCCATATG-GAAGATGCAGGGCGCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACTTGTAGCTCATCGT	XhoI
<b>ORF 134</b>	Forward	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCCTGTGCCAATGCG	XhoI
<b>ORF 135</b>	Forward	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAATACCGCTGAGGATG	XhoI
<b>ORF 136</b>	Forward	CGCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC	BamHI-NheI
	Reverse	CCCGCTCGAG-TTCCGAATATTTGGAACTTTT	XhoI
<b>ORF 137</b>	Forward	CGCGGATCCCATATG-GGCACGGCGGGAAATA	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAACGGTATGCCGCC	XhoI
<b>ORF 138</b>	Forward	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CGGCGTTTTATAGCGG	XhoI
<b>ORF 139</b>	Forward	GCGGATCCCATATG-GCTTTTTTGCGGTAATG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAACGTTTCCGTGCGTTT	XhoI

<b>ORF 140</b>	Forward Reverse	GCGGATCCCATATG-TTGCCACAGGCAGC CCCGCTCGAG-GACGATGGCAAACAGC	BamHI-NdeI XhoI
<b>ORF 141</b>	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTTAAATATT	BamHI-NdeI XhoI
<b>ORF 142</b>	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCGCTCGAG-AAACGTATAGCCTACCT	BamHI-NdeI XhoI
<b>ORF 143</b>	Forward Reverse	GCGGATCCCATATG-GATACCGCTTTGAACCT CCCGCTCGAG-AATGGCTTCCGCAATATG	BamHI-NdeI XhoI
<b>ORF 144</b>	Forward Reverse	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC CCCGCTCGAG-AGATTGTTGTTGTTTTTCG	BamHI-NdeI XhoI
<b>ORF 147</b>	Forward Reverse	GCGGATCCCATATG-TCTGTCTTTCAAACGGC CCCGCTCGAG-TTGTTTTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

**CLAIMS**

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
- 30 5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
7. An antibody which binds to a protein according to any one of claims 4 to 6.
8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,  
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,  
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,  
411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,  
5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,  
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,  
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,  
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,  
611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,  
10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,  
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,  
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,  
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,  
811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,  
15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,  
& 891..

10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the  
group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,  
41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,  
20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,  
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,  
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,  
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,  
255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,  
25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,  
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,  
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,  
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,  
455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,  
30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,  
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,  
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,  
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,  
655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,  
35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

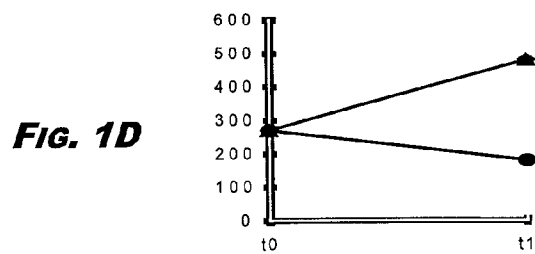
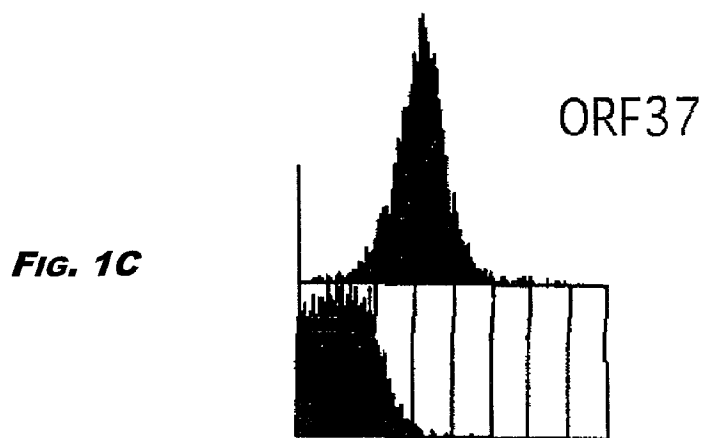
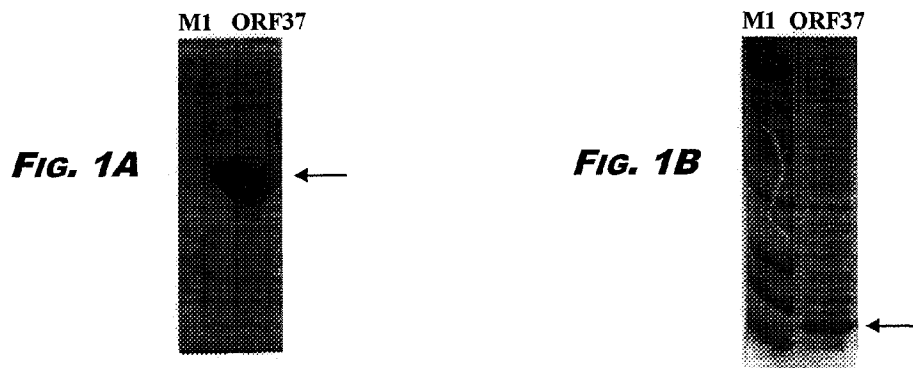
735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any  
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

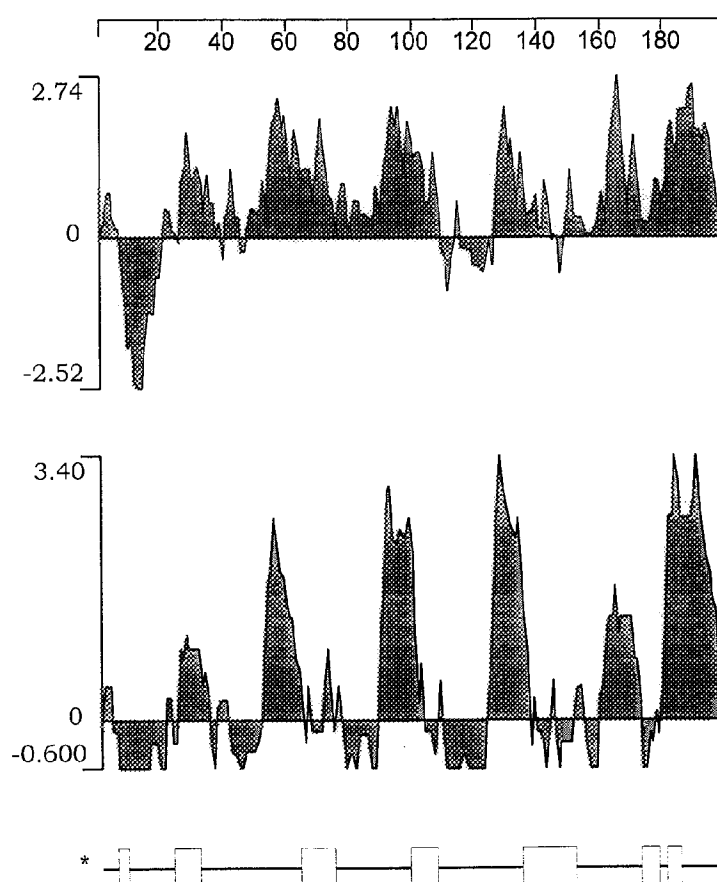


1/24

**FIGURE 1**



2/24

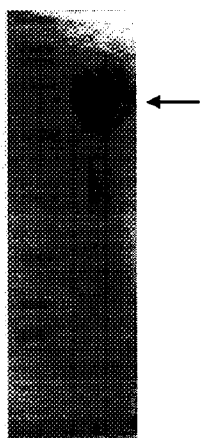
**Fig 1E**

3/24

**FIGURE 2**

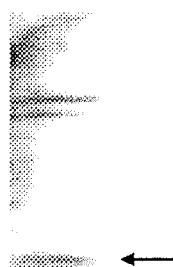
***Fig. 2A***

M1 ORF5



***Fig. 2B***

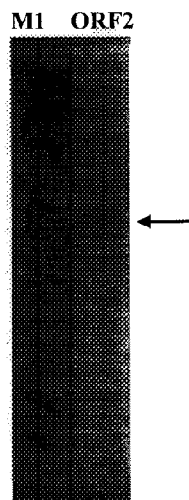
TP



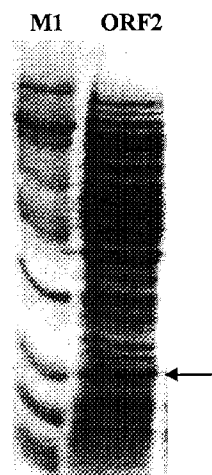
4/24

**FIGURE 3**

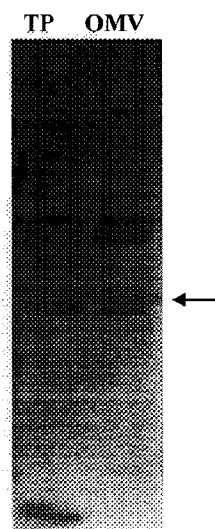
**FIG. 3A**



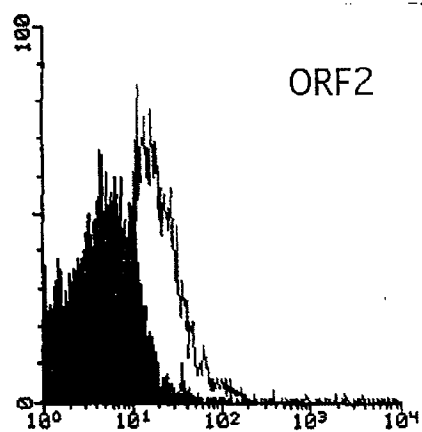
**FIG. 3B**



**FIG. 3C**



**FIG. 3D**

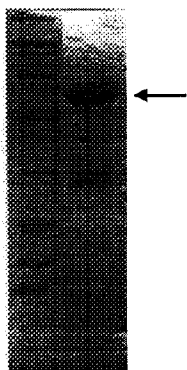


5/24

**FIGURE 4**

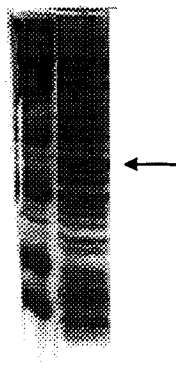
**Fig. 4A**

M1 ORF15



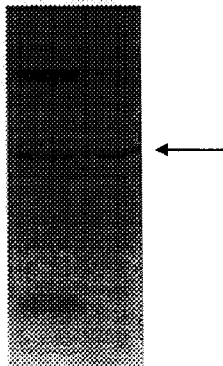
**Fig. 4B**

M2 ORF15



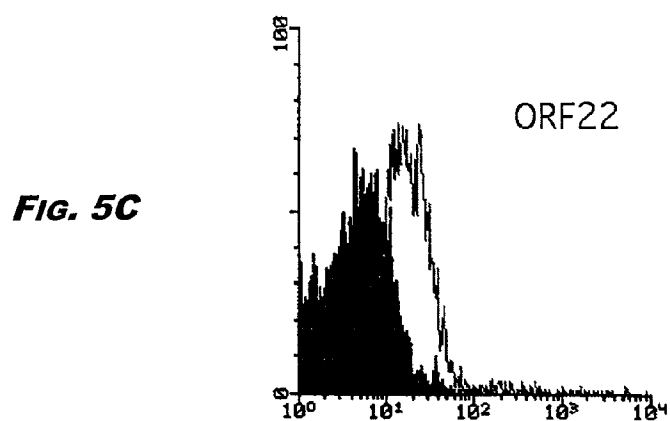
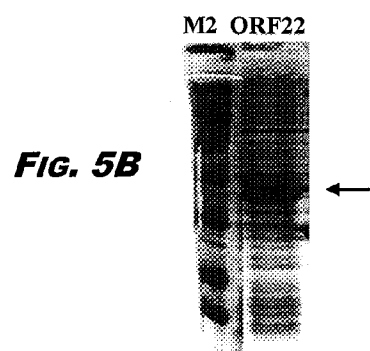
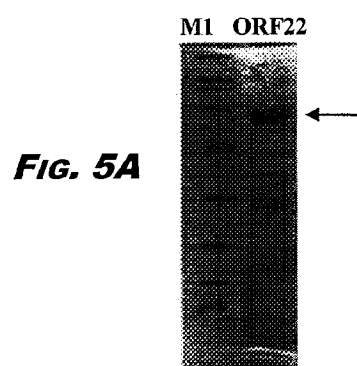
**Fig 4C**

TP OMV



6/24

**FIGURE 5**

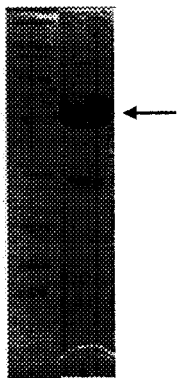


7/24

**FIGURE 6**

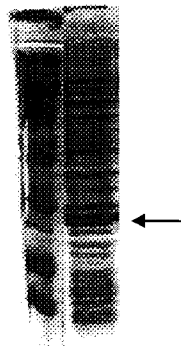
***Fig. 6A***

M1 ORF28



***Fig. 6B***

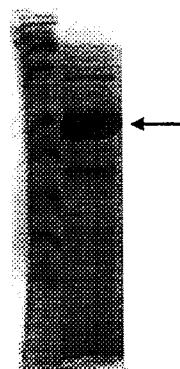
M2 ORF28



**FIGURE 7**

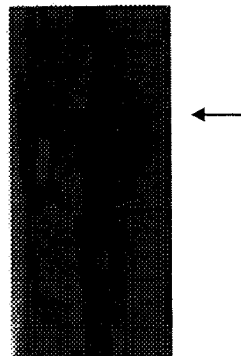
***Fig. 7A***

M1 ORF32



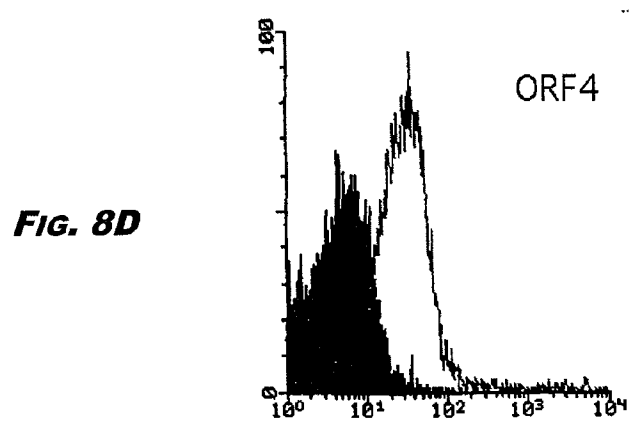
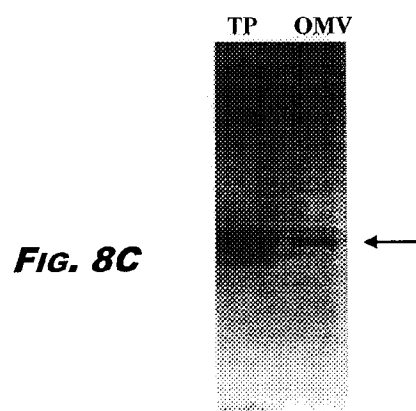
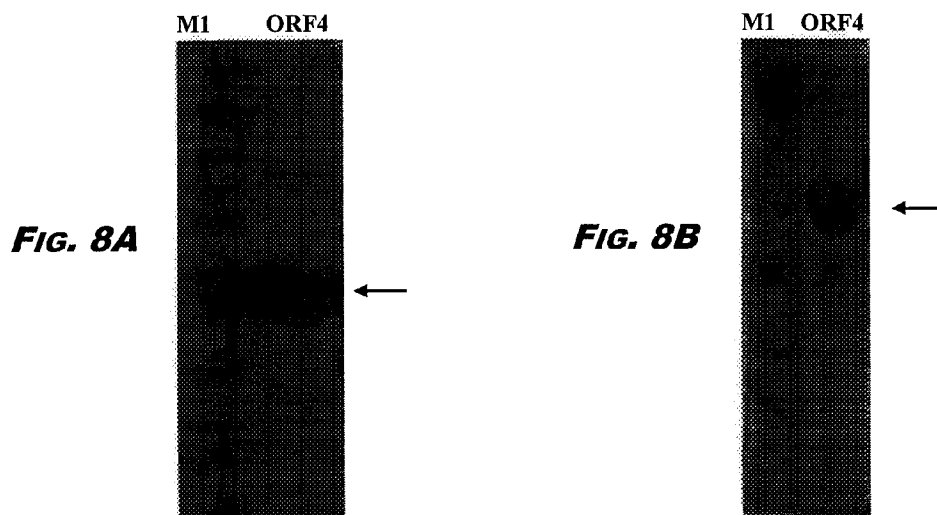
***Fig. 7B***

M1 ORF32



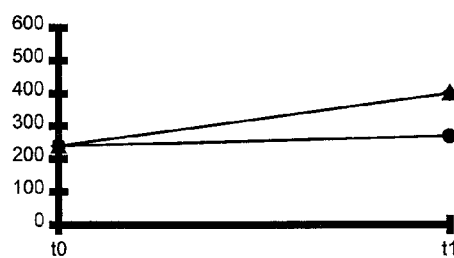
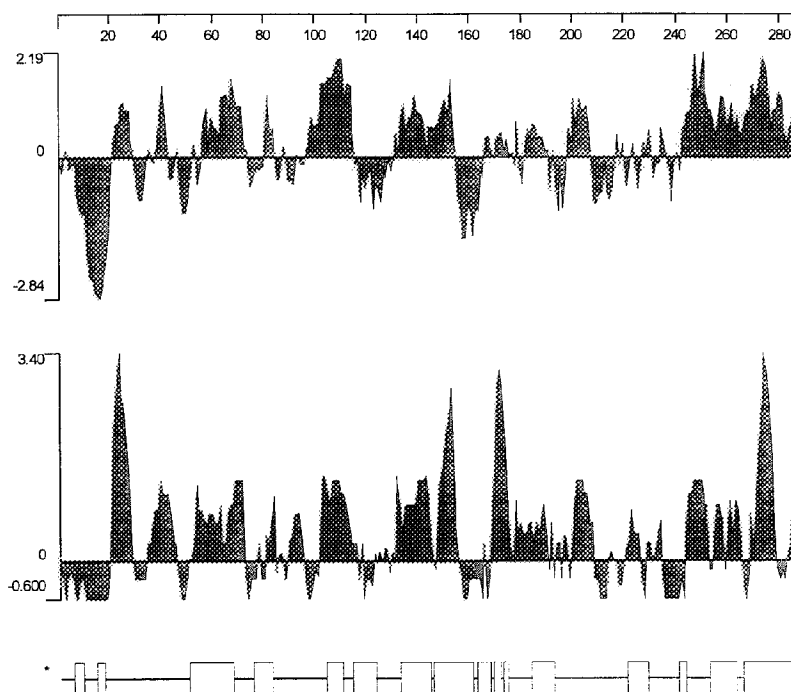
8/24

**FIGURE 8**

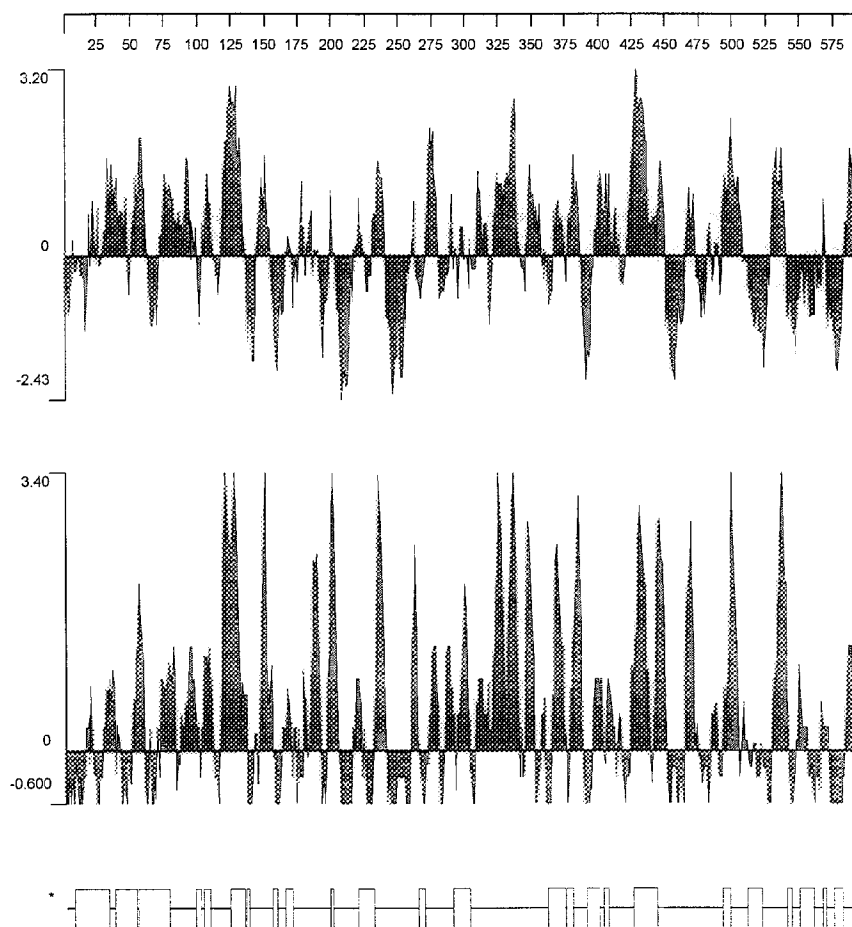




9/24

**FIG. 8E****FIG. 8F**

10/24

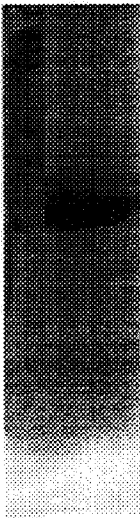
**FIGURE 9**

11/24

**FIGURE 10**

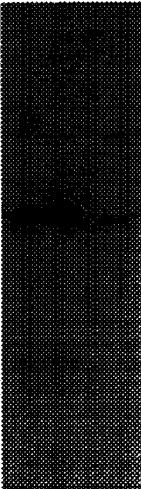
***FIG. 10A***

M1 ORF76

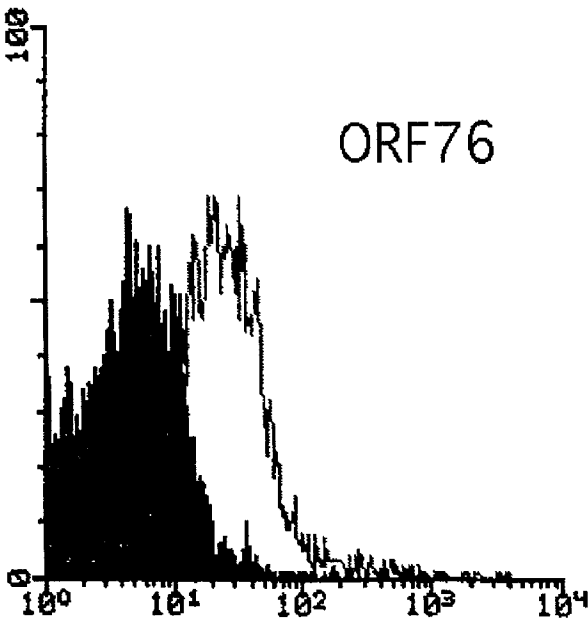


***FIG. 10B***

TP OMV

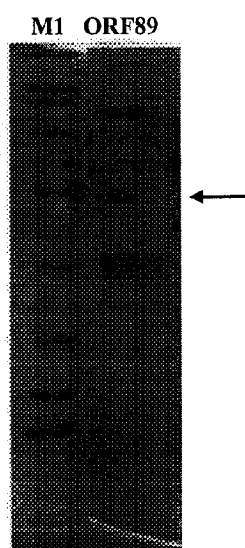


***FIG. 10C***



12/24

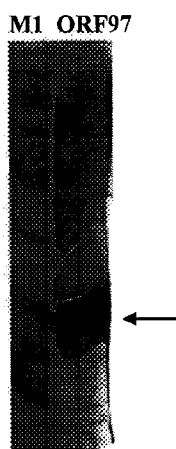
**FIGURE 11**



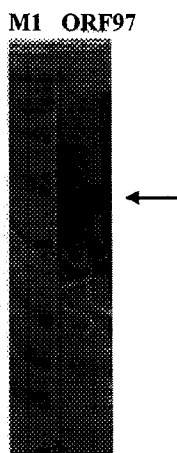
13/24

**FIGURE 12**

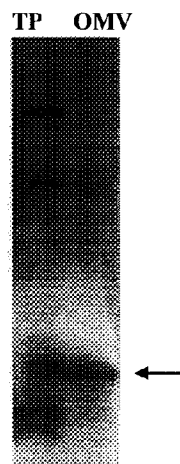
**FIG. 12A**



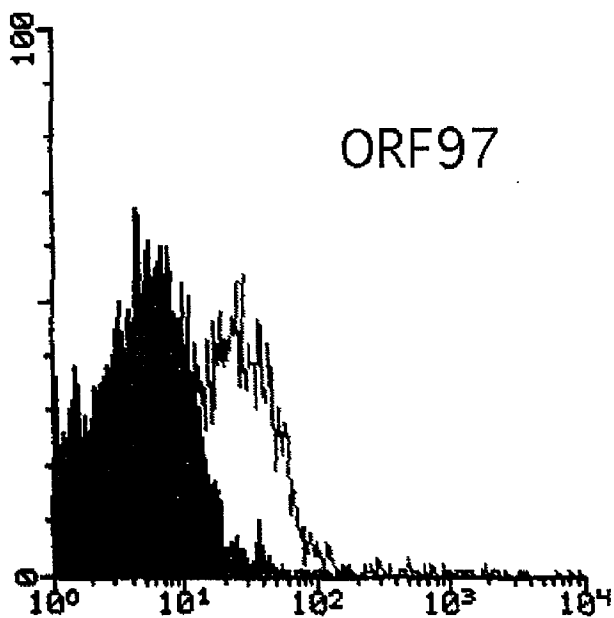
**FIG. 12B**



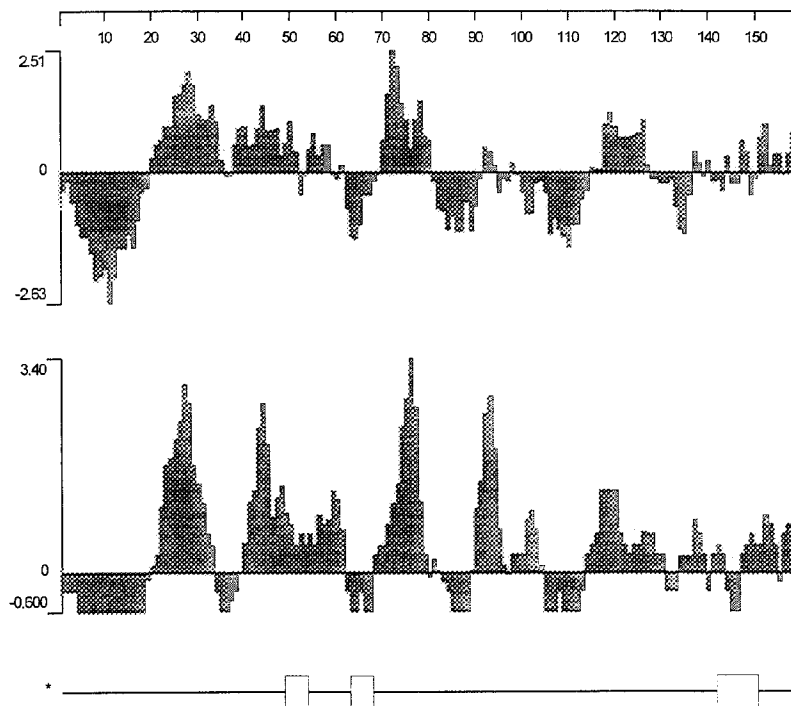
**FIG. 12C**



**FIG. 12D**



14/24

**Fig. 12E**

15/24

**FIGURE 13**

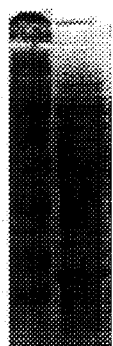
***Fig. 13A***

M1 ORF106

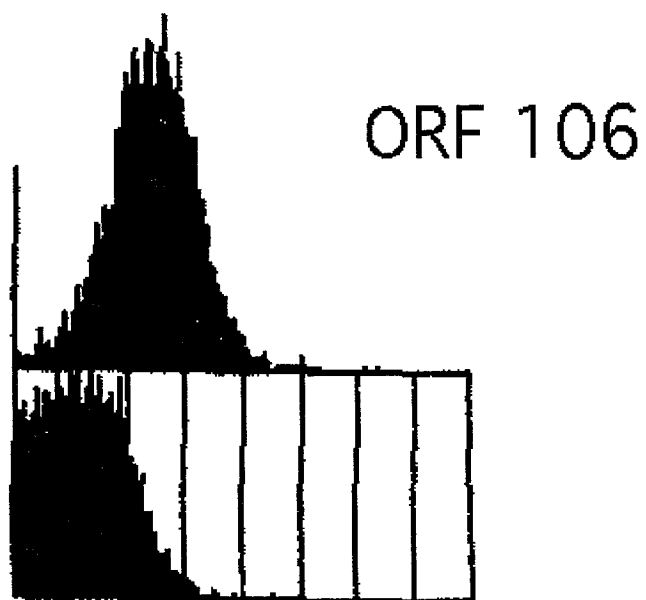


***Fig. 13B***

M2 ORF106



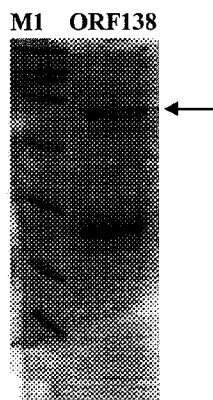
***Fig. 13C***



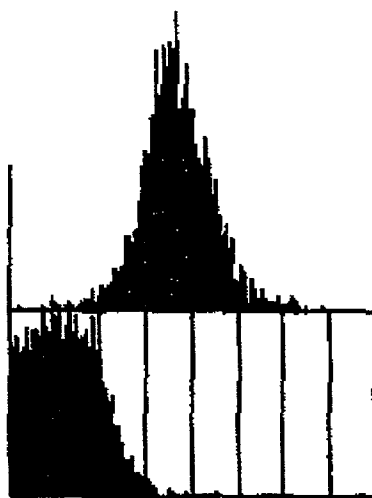
16/24

**FIGURE 14**

***FIG. 14A***



***FIG. 14B***



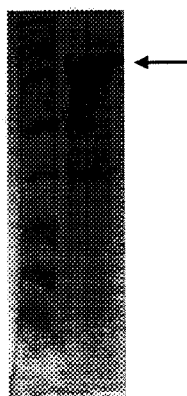


17/24

**FIGURE 15**

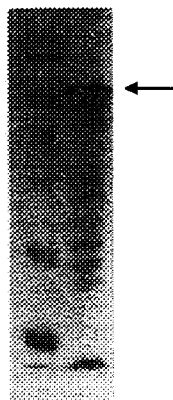
***Fig. 15A***

M1 ORF23



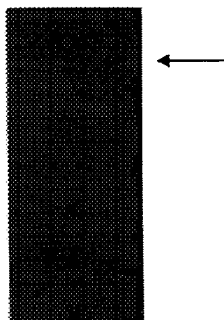
***Fig. 15B***

M2 ORF23

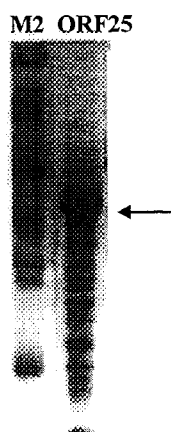
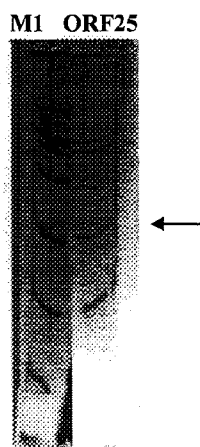
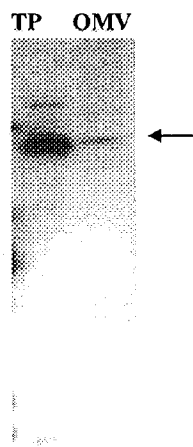
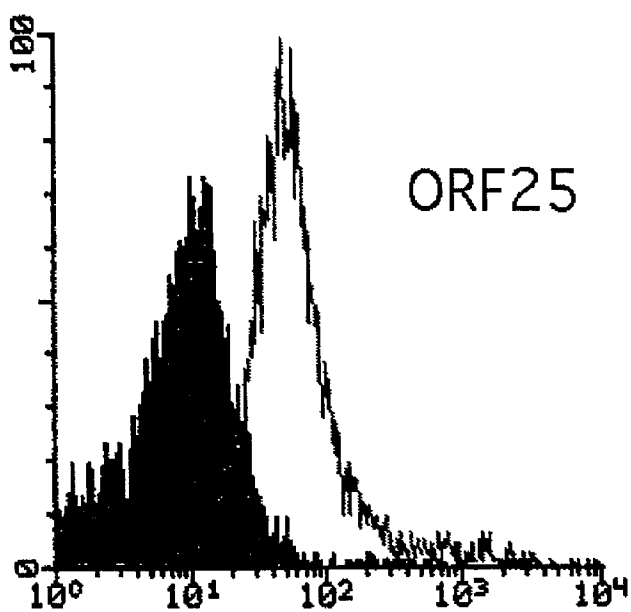


***Fig 15C***

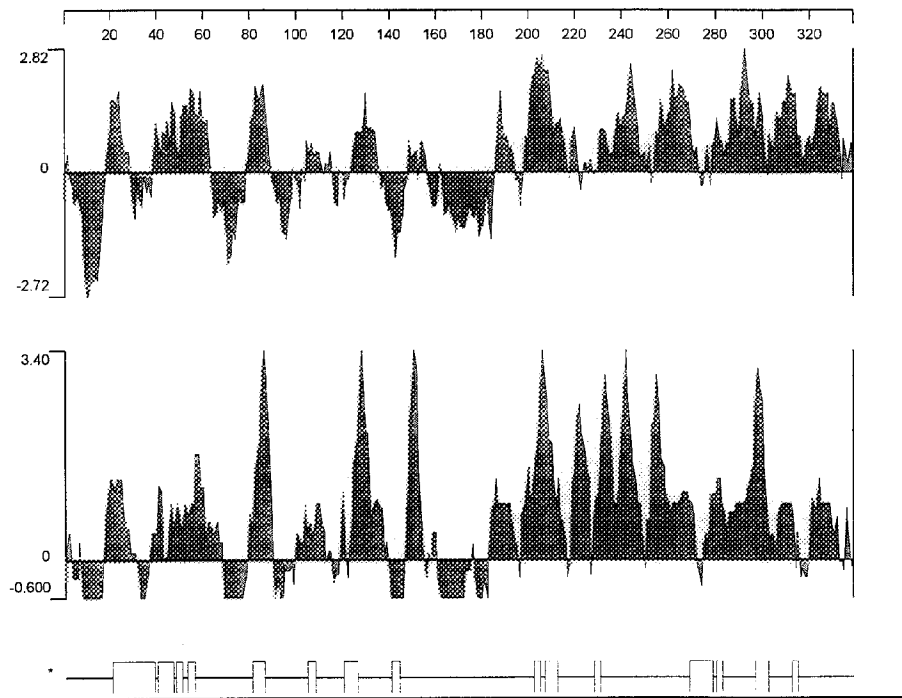
TP OMV



18/24

**FIGURE 16****FIG. 16A****FIG. 16B****FIG. 16C****FIG. 16D**

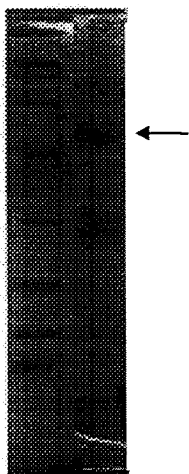
19/24

**Fig. 16E**

**FIGURE 17**

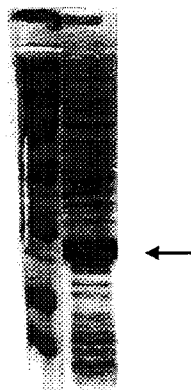
***Fig. 17A***

M1 ORF27



***Fig. 17B***

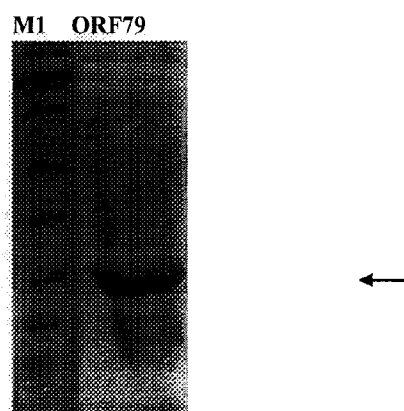
M2 ORF27



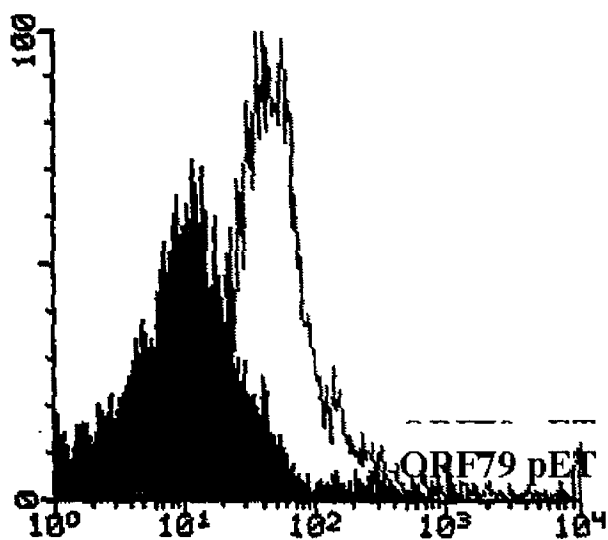
21/24

**FIGURE 18**

***FIG. 18A***



***FIG. 18B***

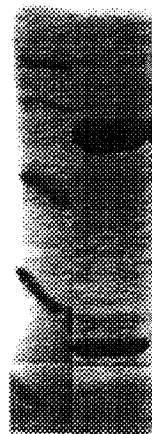


22/24

**FIGURE 19**

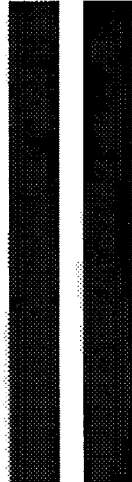
**FIG. 19A**

M1 ORF85

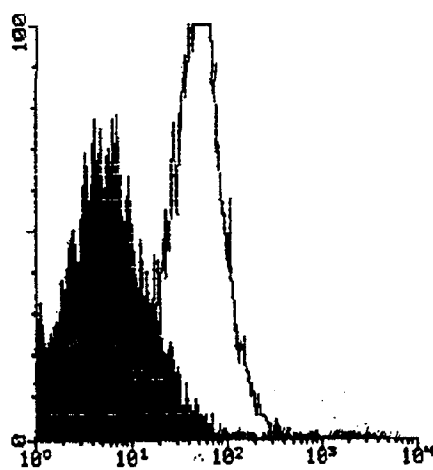


**FIG. 19B**

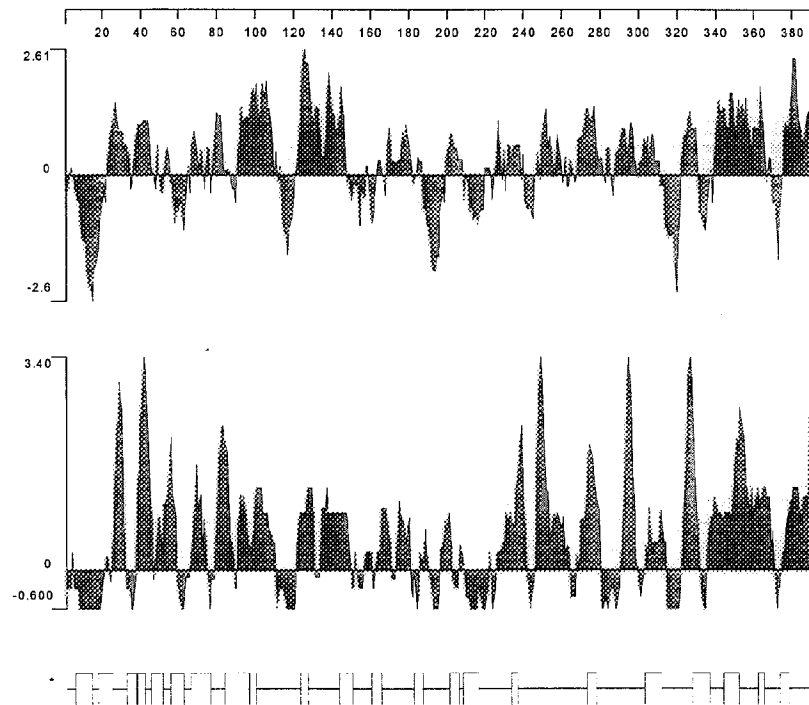
TP OMV



**Fig. 19C**



23/24

**Fig 19D**

24/24

**FIGURE 20**

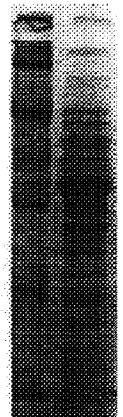
***Fig. 20A***

M1 ORF132



***Fig. 20B***

M2 ORF132



***Fig. 20C***



ORF132